Wearing Many (Social) Hats: How Different are Your Different Social Network Personae?*

Changtao Zhong†1, **Hau-wen Chang**‡2, **Dmytro Karamshuk**§3, **Dongwon Lee**4, **Nishanth Sastry**5

[1] Twitter, UK [2] IBM Watson, US [3] Skyscanner, UK
[4] College of Information Sciences and Technology, The Pennsylvania State University, US
[5] Department of Infromatics, King's College London, UK

[1] czhong@twitter.com [2] changha@us.ibm.com [3] dima.karamshuk@skyscanner.net
[4] dlee@ist.psu.edu [5] nishanth.sastry@kcl.ac.uk

Abstract

This paper investigates when users create profiles in different social networks, whether they are redundant expressions of the same persona, or they are adapted to each platform. Using the personal webpages of 116,998 users on About.me, we identify and extract matched user profiles on several major social networks including Facebook, Twitter, LinkedIn, and Instagram. We find evidence for distinct site-specific norms, such as differences in the language used in the text of the profile self-description, and the kind of picture used as profile image. By learning a model that robustly identifies the platform given a user's profile image (0.657–0.829 AUC) or self-description (0.608–0.847 AUC), we confirm that users do adapt their behaviour to individual platforms in an identifiable and learnable manner. However, different genders and age groups adapt their behaviour differently from each other, and these differences are, in general, consistent across different platforms. We show that differences in social profile construction correspond to differences in how formal or informal the platform is.

Introduction

As social network usage becomes more common, new platforms with large user bases have multiplied, and users are increasingly creating multiple profiles across different platforms. The latest survey from the Pew Research Center reported a sharp rise in multi-platform usage: more than half of all online adults use two or more social networks (Greenwood, Perrin, and Duggan 2016). However, social networks have traditionally been *walled gardens*, and information within a site is not usually exported to other sites. Consequently, using multiple social networks entails the creation of different social network profiles on each platform, and has led to practices such as cross posting of the same information across networks (Lim et al. 2015).

*This work was partially supported by the Space for Sharing (S4S) project (Grant No. ES/M00354X/1). This research was also in part supported by NSF CNS-1422215, NSF IUSE-1525601, and Samsung GRO 2015 awards.

†Work done while at King's College London

‡Work done while at Penn State University

§Work done while at King's College London

In this paper, we are interested in the fundamental act of creating a social network profile on different platforms. Profiles can be seen as a form of "implicit" identity construction (Zhao, Grasmuck, and Martin 2008), and our primary motivation is to see how this identity is negotiated or adapted across different platforms. We attempt to understand profile construction under the framework of Tajfel's Social Identity Theory (Hogg 2006), which suggests that the process of identity creation involves, among other things, *self-categorisation* – defining oneself in relation to other (real or perceived) groups of people, e.g., by following in-group customs, and the consequent *depersonalisation* – the transformation from being an idiosyncratic individual to being a member of a group.

We first attempt to understand self-categorisation in terms of following *platform-specific norms*. Although major social platforms such as Facebook, Twitter, or LinkedIn offer roughly similar social functionalities, there are some feature differences, and more importantly, differences in culture or expected etiquette of acceptable user behaviour (McLaughlin and Vitak 2012; Walther et al. 2008; Cimino 2009). Since some behaviours are considered as inappropriate in particular contexts, perceived norms impose constraints on profile construction (McLaughlin and Vitak 2012; Zhao, Lampe, and Ellison 2016). For instance, Facebook attempts to institute a "real name policy"[1], and some classes of professionals on LinkedIn may face an implicit pressure to wear suits for their profile pictures. An important question, therefore, is whether users follow platform norms and whether this leads to differences in profile construction across platforms.

Although there are platform-wide policies, norms, trends, and etiquettes, profile generation is a deeply personal "explicit act of writing oneself into being in a digital environment and participants must determine how they want to present themselves to those who may view their self-representation or those who they wish might." (boyd 2011). Social network profiles have been seen as "taste performances" – carefully constructed expressions of user taste and personality (Liu 2007). Thus, it is interesting to investigate whether the ways in which profiles are adapted to suit individual platforms are automatically identifiable or whether they are uniquely idiosyncratic. At the same time,

[1]https://www.facebook.com/help/112146705538576

group associations have an important role to play in identity construction (McCall and Simmons 1978), and different audience segments have different expectations about one's public profiles (Rui and Stefanone 2013). Thus, we are interested in understanding how users from different segments (e.g., from different demographics) would express themselves differently inside the same social platform. In summary, we organise our study of identity and profile construction across social networks as follows:

RQ-1 Do users construct their profile identities differently on different platforms?

RQ-2 Are cross-platform differences in profile construction consistent and identifiable?

RQ-3 Do different social groups and demographics have different ways of presenting themselves? Are group-specific aspects of constructing profiles common across platforms?

To answer these questions, and understand the relationship(s) between the different social network personae of a user, we collected a seed set of different profiles of users on up to five social network platforms, by making use of the personal web hosting site About.me[2]. On About.me, users can create personal profiles and provide links to other social platforms. Using these links, we created a matched set of identities on different social networks, all belonging to the *same* user. We focus on the top-4 social networks represented in the dataset – Facebook, LinkedIn, Twitter, and Instagram – as representatives of today's social web. Goffman (Goffman 1959) suggested that the presentation of self involves both the explicit communication – the "given" aspects, as well as implicit, e.g., non-verbal, communication – the "given-off" aspects. To capture both, we draw on both the major aspects of user profiles, which are often used and heavily customised: (1) given: the textual self-description or self-summary of the user on their profile page, and (2) given-off: the profile image itself. Since both the profile image and text can be free form, and individual users can customise almost any aspect, we focus on aspects that are generalisable across social networks.

To the best of our knowledge, this is the first study looking at how user profiles vary across different social networks, both in the aggregate (as collections of user profiles across different networks), as well as in the individual (as a suite of profiles created by the same user across different social platforms). Although this is intended as a first-cut study, we believe this line of research would be crucial to understand, adapt and support the recent rapid rise of multi-platform online social network usage (Greenwood, Perrin, and Duggan 2016).

Preliminaries

Dataset Construction

About.me is a simple social media platform that provides a social directory service. Each user has an individual profile page that includes a short biography and a set of attributes including location, education, achievements, skills, and tags. Importantly, each user can list links to their profiles in other social network sites. As users themselves voluntarily provide such links to their "other" identities, the quality of such links is high and thus ideal for our intended purpose of comparing user profiles across social network sites.

We sampled seed users as follows: First, we searched for random About.me users located in U.S. top-50 cities[3] and collected the tags used by these seed users. We then extended our seed users by adding users associated with at least one of the top-200 most used tags. We crawled About.me profiles of all users that we detected and extracted links to their profiles on other social networks. In total, we have collected 170,348 unique About.me users, of which 116,998 have at least one other social network account listed.

#	Profile image	#	Self-description
0	15,933 (13.6%)	0	2,743 (2.3%)
1	55,447 (47.4%)	1	30,591 (26.1%)
2	36,548 (31.2%)	2	55,598 (47.5%)
3	3,576 (3.1%)	3	24,602 (21%)
4	2,167 (1.9%)	4	3,457 (3%)
5	3,327 (2.8%)	5	N/A
Total	116,998	Total	116,998

Table 1: **Distribution of the number of profiles.** Numbers in brackets indicate the fraction of those users among total number of About.me users.

Analyzing the distribution of links to other social networks that About.me users listed on their profiles, we found that 76% of About.me users in our dataset have linked their profiles to their alternate account in Twitter, whereas users with links to their LinkedIn, Facebook and Instagram profiles took a share of 65%, 46% and 32%, respectively. Based on this, in the following experiments, we focused our analysis around these top-4 social networks. From 116,998 About.me users, next, we have crawled corresponding profiles linked to the above mentioned social networks and extracted the *profile images* and *self-descriptions* of the profiles on each site[4]. Table 1 shows the distribution of the number of profiles with corresponding images and self-descriptions successfully extracted.

Feature Extraction

Profile images: facial features. The state-of-the-art face detection techniques from Computer Vision (Viola and Jones 2004; Wright et al. 2009) can achieve a very high accuracy on various datasets (Jenkins and Burton 2008; Jain and Learned-Miller 2010). Therefore, we adopted two publicly available face detection software to analyse profile images in our dataset: Face++[5] (Zhou et al. 2013) and Mi-

[2]The datasets used in this paper are shared for wider community use at http://nms.kcl.ac.uk/netsys/datasets/aboutme.

[3]While the precise location distribution of About.me users is unknown, searches for users from non-US cities yielded scarce results.

[4]Since Facebook self-description is not publicly available in its new launched API 2.0, we excluded Facebook in our analysis for self-description.

[5]http://www.faceplusplus.com/detection_detect/

	Face++	Microsoft
Faces numbers	$\surd$	$\surd$
Face gender	$\surd$	$\surd$
Face age	$\surd$	$\surd$
Face race	$\surd$	
Face smiling	$\surd$	
Face glasses	$\surd$	
Face pose (3)	$\surd$	$\surd$
Image category		$\surd$
Image color		$\surd$

Table 2: **A total of 11 features, including 3 components of face pose, were provided by Face++ and Microsoft APIs.**

crosoft Face APIs[6]. Each library takes an image as an input and produces a number of facial features (e.g., detected age, smiling score) which characterize the face (or faces) in the image. We summarize these features in Table 2.

Previous studies on Face++ (Jain and Learned-Miller 2010; Bakhshi, Shamma, and Gilbert 2014) have reported a high accuracy for face detection (with $97\% \pm 0.75\%$ overall accuracy) using various manually labeled or crowd-sourced datasets. Our experiments on About.me dataset also suggest a comparable performance from the Microsoft API: we performed a sanity check by testing whether the two APIs are consistent with each other using three facial features common for both APIs – face number, gender and age. Our analyses on face number show that the results from two APIs were consistent for over 90% of images. We next used images with a single face detected and compared the gender and the age of faces as detected by two APIs. Our results suggest that over 88.5% of images are considered to be of the same gender in both APIs, and about 80% of them have less than 10 years difference in the predicted ages. In summary, both APIs are highly consistent with each other.

Profile image: Deep learning Features. We use deep convolution networks (Krizhevsky, Sutskever, and Hinton 2012)–the state-of-the-art approach in object recognition (Deng et al. 2009)–to extract a set of more fine-grained visual features for our machine learning tasks. More specifically, we train a deep convolution network using Caffe library (Jia et al. 2014) based on 1.3 million images annotated with 1,000 ImageNet classes and apply it to classify profile images, and extract two types of visual features: (1) *Deep neural network* features (with $4,096$ dimensions) from the layer right before the final classification layer, which are known for a good performance in semantic clustering of images (Donahue et al. 2013); and (2) *Recognised objects* among the 1,000 Image Object classes that the model is trained on.

Self-description: Word2Vec Features. We analyze the self-descriptions of the users with the Word2Vec methodology proposed in (Mikolov et al. 2013b; 2013a). More specifically, we extract the word vector representation from the corpus consisting of all collected self-descriptions in our dataset using the Gensim natural language processing library[7]. We further reduce the dimensionality of the word space for our analysis as follows. Firstly, we cluster words in Word2Vec dictionary in $N = 100$ groups[8] using k-means clustering algorithm. Each cluster is presumed to represent a group of words located nearby in the Word2Vec space and, therefore, representing different semantics. We then map each self-description in a 100-dimensional feature vector in which each dimension represents the normalized frequency of words from each Word2Vec cluster.

[6]https://www.microsoft.com/cognitive-services

[7]https://radimrehurek.com/gensim/

Male (M)	**Female (F)**	**Total**
63,012	41,197	104,209

(a) Gender

Youth (Y) ≤ 25	**Young Adults (YA)** $26 - 34$	**Adults (A)** ≥ 35	**Total**
11,312	21,059	29,705	62,076

(b) Age

Table 3: **The distribution of user demographics.**

Demographic Inference

Next, we introduce the method we used to infer the demographic information of users (see Table 3).

Gender. To extract reliable gender information, we check the gender information detected from users' profiles images and remove images that are divergent for Face++ and Microsoft APIs.This results in 63,012 male and 41,107 female users. The higher number of males is in line with Alexa traffic statistics for the About.me website[9]. We also validate the performance of the method using data collected from Flickr, which has gender information of 5,627 users. We find that the accuracy of our method is over 98% for males and over 96% for females.

Age. In our age inference, we divide users into three groups, ≤ 25, $26 - 34$ and ≥ 35, to represent Youth, Young Adults, and Adults, respectively, following the age demographic method used in market analysis (International Markets Bureau, Canada 2012). To decide which group a user belongs to, we calculate the average age detected from the user's all profile images, and label the user based on the average age. Removing users that are assigned different age groups by Face++ and Microsoft APIs, 62,076 users remained in our dataset. Then, to validate our method, we applied textual pattern recognition algorithms to parse a list of patterns that specifically describe users age in self-descriptions (e.g., "I am 27 years old", "I'm 23"), adopting Jang *et al.* (Jang et al. 2015). Young Adults are over-represented as compared to US Census Data, but this is in line with increased use of Online Social Networks by younger users[10].

Representativeness of Data

Although we mainly analyze the data from Facebook, Instagram, Twitter and LinkedIn, About.me provided the indispensable link between the identities of a single user across

[8]The value of N has been chosen by optimizing for the classification performance in the machine learning tasks from the later sections.

[9]http://www.alexa.com/siteinfo/about.me

[10]e.g., see https://goo.gl/3aCrVw

the other platforms. About.me was chosen for this study as it was the largest such site at the time of crawling[11], and thus provided the largest breadth and coverage across the major social networks of interest. However, the above differences from 'expected' demographics in terms of age and gender highlight the issue of 'representativeness bias' (Tufekci 2014) introduced by the dataset. Apart from the gender and age bias noted above, according to Alexa siteinfo, About.me also has a more than expected number of visitors with graduate school eduction.

However, the notion of 'expected' can itself be problematic: There can be more than one reasonable baseline – e.g., our data could be compared against the US Census data since our data is primarily from US cities; or it could be compared against all users of social networks. More importantly, Alexa siteinfo indicates that there are differences amongst social networking sites in general – LinkedIn is used disproportionately more by those with graduate school background, while Twitter and Facebook have slightly more than average number of visitors with an educational background of 'college' and 'some college' respectively, and Instagram user base is more female than male. This raises the interesting question of what 'representative' means for a multi-platform study such as ours: A user who has an account on Instagram may not be representative of users on LinkedIn and vice-versa. Indeed, users who actively use (or have accounts on) multiple social networks are likely not representative of the 'average' social network user – thus a user who has an account on both Instagram and LinkedIn is unlikely to be a representative user of either platform, or of an entirely different platform such as Facebook.

Yet, given the prevalence of and recent increase in the use of multiple social networks (Greenwood, Perrin, and Duggan 2016; Zhong et al. 2014), we expect that effects such as the ones we see here may become more and more typical, even if it is not typical usage today, and therefore believe our results are a valuable insight into how users shape their online personae across social network platforms. Follow-on work may be required to confirm the wider applicability of our results, and to better understand whether these kinds of bias may be an unavoidable issue for any multi-platform study, or if there are ways to overcome some of them.

RQ1 - Different Norms on Different Networks?

In the following, we aim to answer RQ1 and quantify differences between the ways About.me users present themselves on different social platforms. Informed by Goffman's self-representation theory (Goffman 1959), we search for both explicit verbal self-expressions they attach to their profile texts as well as implicit hidden cues they encode (whether intentionally or not) in their profiles images. For instance, we examine whether a profile picture features friend(s), sunglasses or an outdoor landscape, or whether it contains a smiling portrait picture. We interpret these as indicators of norms of formality or informality – or alternately, the extent to which a user is managing her public impression[12].

Face Number. In Figure 1a, we analyze the number of faces detected on the profile images across multiple social networks. Note that on LinkedIn, presumably a more formal social network, the majority (90%) of profile images are portraits of a single person, while Facebook and Instagram, presumably more informal social networks, less than 60% of images are portraits of a single person. Similarly, both Twitter and About.me have a higher proportion of profiles that have portraits than Facebook and Instagram. The latter two platforms appear to surface an interesting convention: a significant minority (up to 40%) of users who use a non-face image (e.g., cartoon, outdoor landscapes, etc.) as their profile picture. In alignment with this convention, Facebook has also a non-trivial share (15%) of group portraits with more than one face (unlike LinkedIn profiles where this share is less than 1%), which may be attributed to the emphasis that Facebook users place on intimate social relationships in their profiles (Mod 2010; Strano 2008).

Smiling Score. Next, we detect the extent to which users are smiling on their profile images by measuring the so-called *smiling score*, ranging from 0 (i.e., no smile) to 100 (i.e., laugh) as provided by Face++ API. Note that only profiles with 1 face are considered here. Figure 1b presents the distribution of smiling scores in user profile images for 5 social networks. The profile images on more formal platforms (e.g., LinkedIn) tend to have higher smiling scores than those on informal ones (e.g., Instagram and Facebook). We take this is an indication that users tend to manage their professional impression on more formal social platforms – such as LinkedIn than they do in informal ones – such as Facebook.

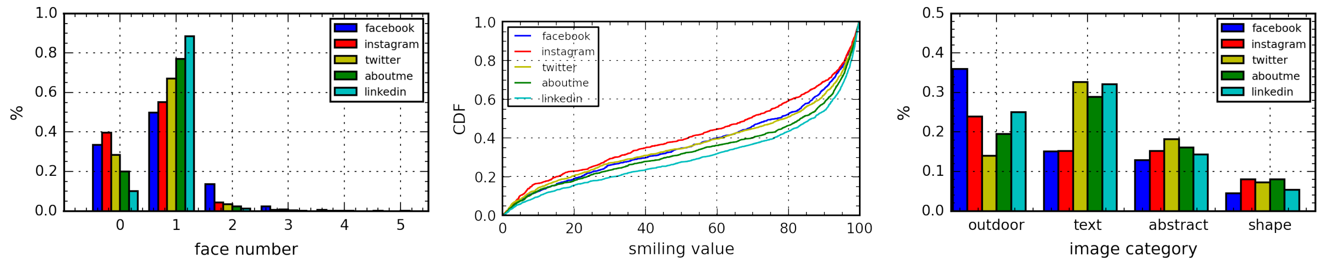
Eyeglasses. From Face++ API, we could identify two types of eyeglasses: normal eyeglasses and sunglasses. For normal eyeglasses, we find a consistent trend across all social networks with about 17% wearing them. But sunglasses, which usually are related to leisure activities, are less likely to appear in profile images from formal platforms (e.g., 2.1% in LinkedIn), compared with those from informal social platforms (e.g., 7.3% in Facebook and 7.2% in Instagram).

Image Category. Finally, we examine the categories of profile images detected by the Microsoft API[13]. To this end, we exclude all portraits from our analysis ($\approx$ 50% of all profile images) and analyze the image categories of all remaining profile images. Figure 1c presents the distribution of the top-4 image categories, namely, *outdoor*, *text*, *abstract*, and *shape*, each of which accounts for at least 5% of non-facial

[11]When the data was collected in Aug 2015, About.me was ranked in the top 3000 of all websites in the Alexa rankings, and although it is ranked ≈ 5000 now, it still remains the largest such website.

[12]To recognise people and faces, we have experimented with both Face++ and Microsoft APIs for detecting the number of faces and have seen a high level of consistency between two. However, the results are similar, and we only present the results from Face++. In contrast, our analyses on smiling scores and glasses are based only on Face++ API, while image category results only on Microsoft API, since these features were only available in one of the libraries.

[13]The taxonomy of image categories can be found at https://www.projectoxford.ai/doc/vision/visual-features.



(a) Face number (b) Smiling score (c) Image category

Figure 1: **The distribution of the detected (a) face number, (b) smiling score, (c) image category of users' profile images at different social networks.** Note that we only consider images with 1 face in (b) and images not in the "people" category in (c).

images in every social network. We observe that Facebook and Instagram users tend to use more of *outdoor* images on their profiles. This can be explained by the emphasis on non-professional activities – such as travel experience – in communicating with their peers on these networks (Sharda 2009). In contrast, *text*, *abstract*, and *shape* images are more dominating among Twitter and About.me profiles. A non-exhaustive manual inspection suggests that this is, in some cases at least, an instance of expressing themselves as recognisable brands.

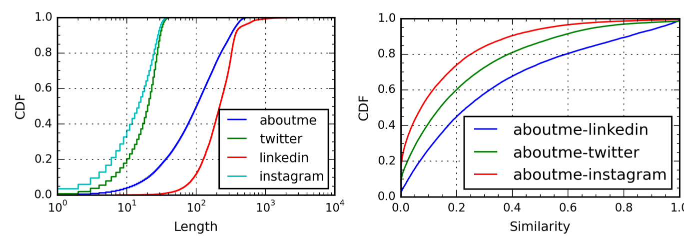
Differences in Self-descriptions. Due to the variation of functionality and specific limitations of social networks, a user might tailor her profiles differently for a given platform. As Figure 2a shows the distributions on the length (i.e., the number of words) of self-descriptions across platforms, note that both About.me and LinkedIn are skewed to the right, indicating that most of self-descriptions are relatively long. However, the distributions of Twitter and Instagram are somewhat irregular, probably due to the artificially-set length limitation.

Next, for each possible pair of social network profiles of a user, we compute the TF-IDF similarities between the self-descriptions[14]. The features are normalized to reduce the impact of profile length. The similarity score is between 0 (i.e., very different) and 1 (i.e., exact match). For instance, demonstrated in Figure 2b is the CDF of similarity scores for About.me. As expected, the self-description of About.me is most similar to that of LinkedIn for the same user, presumably due to the relatively similar functions of both social networks.

Word Clouds. As discussed in previous sections, each social network tends to develop its own culture, which results in a variation of profiles even for the same user. To summarize the themes of each social network, in Figure 3, we visualize self-descriptions in profiles as word clouds, generated by a Python library[15]. It is interesting to see that LinkedIn and About.me self-descriptions concern topics around employments, including professional terminologies (e.g., development, project, experience), types of industry (e.g., marketing, media, business, management), location (e.g., "new"

[14]Recall that there are no self-descriptions from Facebook.

[15]https://github.com/amueller/word_cloud



(a) Lengths of self-description (b) Similarity (About.me)

Figure 2: **Difference of self-descriptions**: (a) Distribution of lengths of self-description and (b) pair-wise TF-IDF similarity scores across different social networks.

for location names like New York and New Jersey) and experiences (e.g., "year"). However, Instagram self-descriptions show more relaxed roles of users such as "life", "love", "lover", "food", "music", and "travel". On the other hand, Twitter self-descriptions are somewhat a mix between two groups, heavily comprised of words such as "love", "marketing', "write", and "social".

Discussion. The above results suggest that there is a spectrum of norms across the social platforms, with a consistent difference between the more *professional* networks such as LinkedIn and the more *informal* networks like Instagram (Van Dijck 2013). Many if not all of the above differences can be understood in terms of the professional or non-professional nature of the platform.

The fact that a single user may have more than one kind of profile is explained by the faceted identity theory (Farnham and Churchill 2011), which posits that people have multi-faceted identities, and enable different aspects of their personalities depending on the social context. Users may be more focused on managing public impression (Roberts 2005) in professional circles, but may not be when with friends and family. The strength of the boundaries between these different social roles may vary across individuals (Clark 2000; Farnham and Churchill 2011) and may impact the patterns of daily routines (Ashforth, Kreiner, and Fugate 2000) to different extents for different individuals.

(a) About.me (b) LinkedIn (c) Twitter (d) Instagram

Figure 3: **Word clouds made from self-descriptions.**

More interestingly, there are empirical evidences (Ozenc and Farnham 2011) suggesting that, to manage communication within different social roles, people strategically employ different social media channels. For instance, on social networks focused on building and maintaining external professional networks (e.g., LinkedIn) – users may want to present themselves in a formal and professional way – whereas on general-purpose social networks (e.g., Facebook), users may keep relationships more informal and adjust their profiles accordingly (Skeels and Grudin 2009).

RQ2 - Profile Classification

Building on the analysis from RQ1, now, we ask whether users fit in to the norms of those social networks and attempt to answer this indirectly by studying how accurately one can predict profiles in different social networks. Note that we build a non user-specific model to see whether users fit in to a given social network in a consistent manner . We formulate the *profile classification* problem as follows:

PROBLEM 1 (PROFILE CLASSIFICATION) *Given a profile image (or self-description) of a user u, predict whether it* fits in *the profile conventions of a given social network n.*

We tackle this problem using a supervised learning approach where we exploit profile images or self-descriptions originated from different social networks and train a classification model to predict the social networks that they have been picked from.

For each profile image in our dataset, we extract the aforementioned facial and deep learning features, including face number, gender, age, race, smiling, glasses, and face pose. We also consider 5,096 features extracted using convolution neural networks including a 1,000 dimensional vector of recognized objects and a 4,096 dimensional vector extracted from the pre-trained deep neural network. Similarly, we also extract textual features from self-descriptions and construct a Word2Vec vector. For the purpose of this analysis, we validate the models using a Random Decision Forest classifier[16] and report the results of the 10-fold cross-validation.

[16]We used a Random Forest implementation from the SKLearn package with $\sqrt{n_{features}}$ split and 100 estimators (other values from 10 to 1000 were also tested, but 100 showed the best trade-off between speed and prediction performance). χ^2 feature selection is applied to select 1000 most relevant features for the classification with profile images.

	ACC	**P**	**R**	**F**	**AUC**
Instagram	0.829	0.790	0.896	0.840	0.829
Facebook	0.770	0.746	0.824	0.782	0.769
About.me	0.730	0.723	0.748	0.736	0.730
LinkedIn	0.687	0.698	0.665	0.681	0.687
Twitter	0.657	0.644	0.707	0.673	0.657

(a) Using profile images

	ACC	**P**	**R**	**F**	**AUC**
Instagram	0.768	0.571	0.288	0.383	0.608
About.me	0.866	0.765	0.643	0.716	0.802
LinkedIn	0.871	0.720	0.797	0.756	0.847
Twitter	0.788	0.598	0.468	0.525	0.681

(b) Using self-description

Table 4: **Performance of the *one-vs-others* profile classification.** We evaluate the performance using Accuracy (ACC), Precision (P), Recall (R), F-score (F), and Area Under the receiver-operator characteristic Curve (AUC).

One-vs-others Classification. To start with, we train a set of one-vs-others classifiers (one for each social network) which for a given image or self-description are trained to distinguish between social networks that they belong to. To this end, for each classifier, we label profile images/self-descriptions picked from the corresponding social network as positive instances and randomly sample the same number of profile images/self-descriptions from the other four social networks (i.e., "others") as negative instances. The results of the one-vs-others experiments using profile images and self-descriptions are summarized in Table 4.

First, we note a high prediction performance for one-vs-others profile classification problem–e.g., high AUC scores of up to 0.829 for Instagram profile images and up to 0.847 for LinkedIn self-descriptions. This suggests that profile conventions of individual social networks can be successfully recognised by machines with high accuracies. However, prediction performance varies a lot across social networks. For example, the AUC of classifying Twitter profiles is only 0.657 in comparison with the best-in-class AUC of 0.829 of Instagram. In general, we observe that profile conventions in informal social networks (e.g., Facebook and Instagram) are much more recognisable using profile images than in professional social networks (e.g., LinkedIn and About.me). In contrast, self-descriptions in professional social networks perform much better than in informal social networks. This implies that the conventions of professional social networks

	Instagram	Facebook	About.me	LinkedIn
Facebook	0.744			
About.me	0.905	0.873		
LinkedIn	0.890	0.834	0.691	
Twitter	0.879	0.835	0.571	0.599

(a) Using profile images

	Instagram	About.me	LinkedIn
About.me	0.894		
LinkedIn	0.863	0.869	
Twitter	0.665	0.867	0.974

(b) Using self-description

	Instagram	About.me	LinkedIn
About.me	0.952		
LinkedIn	0.904	0.756	
Twitter	0.880	0.862	0.639

(c) Using combined features

Table 5: **The *One-vs-one* performance (AUC) of profile classification.** Note the emergence of two groups, separated by dotted lines. Intra-group prediction is low while inter-group prediction is high.

are mainly expressed through users' self-descriptions, while profile image is a channel for expression on informal social networks.

One-vs-one Classification. In the *one-vs-one* classification, we aim to identify profile images between pairs of social networks: i.e., given profiles from *two* different social networks, we train a binary classifier to distinguish an origin social network for each given profile. The results of the experiments are summarized in Table 5. We note that the results highlight two distinct groups, columns in Table 5 separated by dotted lines, among considered social networks–Instagram and Facebook, on the one hand, and About.me and LinkedIn, on the other hand, with Twitter in between two groups. In general, intra-group prediction is low while inter-group prediction is high. For instance, AUC for the Instagram-Facebook pair is 0.744, while that for the Instagram-About.me pair is 0.905.

In other words, this result highlights that the conventions inside each of the two groups – professional networks (i.e., About.me, LinkedIn and to some extent Twitter) on the one hand, and the more informal networks on the other hand (i.e., Instagram and Facebook) are very similar. This intra-group similarity makes it difficult to distinguish between profile images randomly picked from two. Similarly, it is much easier to distinguish between a profile image taken from a professional social network and an informal one, suggesting that conventions among the two groups of networks are very different. We note that this result resonates with our findings from previous sections.

In Table 5c, finally, we use the combined hybrid features to *one-vs-one* experiments. It shows that the combined model takes the advantages of both profile images and self-descriptions, and improves most pairwise predictions from Table 5b, with the exceptions of About.me-LinkedIn and Twitter-LinkedIn pairs.

Summary. In this section, we demonstrated that using either profile images or self-descriptions, it is indeed possible to accurately predict profiles in different social networks. Indeed, it shows that most of users tend to fit in (by means of profiles) to the conventions of a particular social network. In addition, the proposed profile classification problem has a practical ramification. For instance, by turning the classification problem into the recommendation problem, one can build a tool to recommend the most appropriate profile image (among many choices) on a particular social network site for a given user.

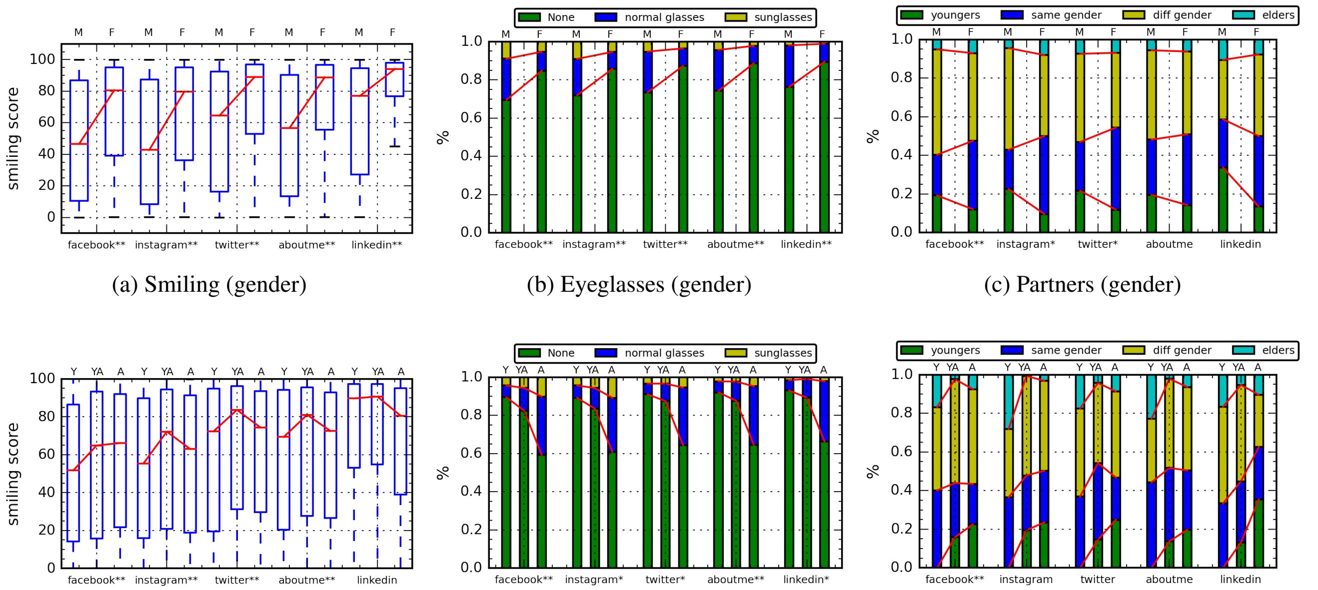
RQ3 - Gender and Age Differences?

In the previous two sections, we have discussed conventions found in user profiles across multiple social networks and analyzed the extent to which user profiles *fit in* the conventions of individual social networks. Building on these analysis, in the current section, we further investigate the differences between the way users from different demographic groups express themselves in their social media profiles. To this end, we divide users into different groups according to their gender and age information and analyze the differences in profile images and self-descriptions across different groups. In particular, we examine the discrepancies in aspects such as *smiling score*, *eyeglasses* and *partners* as identified by the facial recognition libraries.

Smiling Scores. We first examine emotional expression in profile faces. To do so, we focus on profile images with a single face only, and compare the smiling scores across different demographic groups. In Figure 4a, we compare the distributions of smiling scores for users with different genders. We note that women tend to smile more in the profile images across all five social networks (with $p < 0.005$). This result is consistent with the several previous studies in psychology (Coats and Feldman 1996; McClure 2000; Strano 2008; Tifferet and Vilnai-Yavetz 2014). Indeed, according to Coats and Feldmen (Coats and Feldman 1996), women who display positive emotional expressions tend to be rated of a higher social status, while men who do so risk being rated as having low social status; thus there is motivation for women to fit in by smiling, and for men not to.

We also compare the smiling score for users in different age groups in Figure 4d. On the one hand, we find that users in adult group (A, age ≥ 35) tend to have lower smiling scores than young adult group (YA, age in $26 - 34$) ($p < 0.005$, except LinkedIn), which is consistent with existing psychological theories (Fölster, Hess, and Werheid 2014; Gross et al. 1997) that older people tend to appear less expressive. On the other hand, surprisingly, we find that users from the youths group (Y, age ≤ 25) also have lower smiling score. We suspect this is due to higher popularity of novel image categories such as selfies amongst youth, where smiles are less common or less pronounced (Souza et al. 2015).

Eyeglasses. We also examine eyeglasses people wear in profile images. As discussed before, we identify two types of eyeglasses: normal glasses and sunglasses. Normal glasses are mainly used for vision correction. In Figure 4b (blue



(a) Smiling (gender) (b) Eyeglasses (gender) (c) Partners (gender)

(d) Smiling (age) (e) Eyeglasses (age) (f) Partners (age)

Figure 4: **The comparison of the facial features with different genders and ages.** For gender analyzes (a-c), "M" represents results of males, and "F" is for females. For age analyzes (d-f), "Y" is for youth, "YA" is for young adults and "A" is for adults. We examine the differences among distributions for different genders/ages using Kolmogorov-Smirnov test, and label "**" for cases with p-value $p < 0.005$, and "*" for cases with $p < 0.05$.

parts), we compare the usage of normal glasses for different genders. Interestingly, we see that more males wear normal glasses than females (with $p < 0.005$), especially for platforms like LinkedIn. This result is in contrast with the findings of the U.S. National Eye Institute (NEI) (National Eye Institute, U.S. 2010) on myopia, the most cause for corrective eye lenses (National Eye Institute 2010): the NEI report finds that women have higher myopia rate than men. A potential explanation is that wearing normal glasses is considered more intelligent and formal (Edwards 1987). We see more normal glasses for males, because males are thought to dress more formally than females (Chawla, Khan, and Cornell 1992; Sebastian and Bristow 2008). An alternative explanation can be that there is a social pressure on women not to wear glasses, which may influence their choice of profile image. Both explanations, however, are indicative of a choice towards fitting in with expected gender-specific norms or social pressures.

Partners. So far, we have studied how users look like in their profile images. We next explore the partners that appear in users' profile images. We focus on profile images with 2 faces[17]. Since for each user we have several images, we could easily identify the user (u) and the partner (p) from 2 faces by matching gender and age. Then we decide the relationship between them. To do this, we check the age differences between u and p with a threshold X[18]. If $|age(u) - age(p)| > X$, we consider they are in different generations, otherwise in the same generation. When u and p are in different generations, we compare their ages again, if $age(u) > age(p)$, we say the image is with "younger", otherwise, we say with "elder". When u and p are in same generation, we compare their genders, if $gender(u)$ and $gender(p)$ are the same, we say the profile image is with "same gender" friends, otherwise, we say it is with "different gender" friends.

Using this framework, we compare partners for users in different groups. In Figure 4f, we find that with increasing age, users tend to have more profile images with "youngers", which is likely be their kids, and less images with "same gender" friends. For "elders", or parents in most cases, users in $26-34$ group have more images then users in ≤ 25 and ≥ 35 groups. Surprisingly, in Figure 4c, we observe males have more images with "youngers" (likely, kids) and "different gender" friends than females. But females tend to use more images with "same gender" friends. This result is consistent with (Strano 2008), which finds that females are more likely to emphasize friendships in their profile images than males.

Summary and discussion In this section, we have answered RQ3, showing that although users are adapting to social norms of a given platform, there are distinct differences in the way that different genders and age groups present themselves. For example, females smile more than males and tend to avoid the use of spectacles in their profile pic-

[17] Due to the data limit, as shown in Figure 4c and 4f, the results of partners are only statistically significant for Facebook with $p < 0.005$, although similar trends can also observed in other networks.

[18] We present results with $X = 20$, although similar results can be observed for $X = \{10, 15, 25, 30\}$

tures. Although these differences are interesting in and of themselves, it is more interesting that there are statistically significant and consistent differences across the population, indicating that profile construction is informed by gender and age demographics. Gender and age are the most basic social groups. It would be interesting to consider whether other more sophisticated groupings of people (e.g., by party affiliation) tend to construct their profiles in similar ways within the group.

Discussion, Limitations, and Conclusion

In this paper, we used the personal web pages of 116,998 users from About.me to identify and extract a matched set of user profiles on four major social networks (Facebook, Twitter, ing and LinkedIn), and studied the following three questions: (1) whether users express themselves differently on different social networks, to "fit into" the "culture" or expected "norms" of the network (2) whether different users "fit-in" in a consistent manner, i.e., whether we can learn a non user-specific model that distinguishes a profile as fitting or belonging to one social network more than another, and (3) whether there are any universal cross-platform tendencies in how various genders and age groups fit into a given platform.

Answers to these questions can have deep implications for core applications such as advertising or community building: The extent to which norms exist as well as the extent to which users (need to) adapt to fit those norms influences what kinds of users are attracted to each network. This can affect advertising strategies (e.g., brands and advertising styles that find a large audience on say LinkedIn may not be popular on Instagram, because the expected norms and core demographic may be different). This approach can also facilitate automated generation of stereotype personas - fictitious individuals representing segments of the audience in marketing, design and advertising (An, Kwak, and Jansen 2017).

Similarly different strategies may be needed on different platforms for community engagement and growth (e.g., gamification and rewards may promote community engagement on one social network which attracts one kind of user, but not on another network with a different user base that does not like the competitiveness of gamification). If users "fit in" in consistent patterns on a platform, then new UI affordances can be developed to help them fit in. At the same time, if users in different demographic groups have different ways of fitting in, or feel pulled to fit in to different extents, then UI affordances need to be sensitive to and support such differences.

Our results indicate that different social networks do have different conventions, and users do fit their different social networks profiles to suit these conventions. Importantly, we confirm that the networks we examine – Facebook, Instagram, Twitter, LinkedIn and About.me – fall on a spectrum, based on how formal or informal the network's intended purpose is. Profiles on more formal or professional social networks such as LinkedIn and About.me are presumably intended to be showcased to an audience of non-friends (e.g., recruiters on LinkedIn, or random visitors to personal webpages on About.me) and are therefore constructed differently from profiles on networks such as Facebook or Instagram, which appear to be geared more towards friends and relationships. We find these differences both in conventions relating to profile images (e.g., usage of pictures with friends and companions on Facebook as opposed to formal single-person portraits on LinkedIn) as well as the choice of text in profile descriptions (Figure 3).

Users' fitting in to conventions are recognisable by simple machine learning classifiers; networks that are farther apart on the formal / informal spectrum are easier to tell apart from each other. We also find evidence that different genders and age groups fit in to different extents, and that the relative extent of fitting in, in general, is consistent across social network platforms. For instance, women are more likely to have pictures with companions and without glasses. While these differences are important in and of themselves (and we try to explain how these differences arise), in the context of this paper, it is more interesting that such differences exist at all, and that are statistically significant and consistent across the platforms. This allows us to draw conclusions on how social network profiles vary based on how formal or informal the social network platform is.

This paper is intended as a first look at how user profiles vary in the aggregate across different social networks. Although we observe robust effects, as noted before (in the Representativeness of Data subsection), our results may be coloured by the fact that our data is limited to users who have made a profile on About.me. A second limitation of our study is that it is based solely on data analysis. It would be interesting to validate whether our conclusions correspond to user motivations in choosing a particular profile image or writing a particular textual self-description. This requires a direct user study, which was out of scope for the current work.

Our choice of relying solely on data analysis was driven by the observation that even if users *believed* they were not trying to fit in, the data might indicate an *unconscious* bias in user choices towards fitting in with a (potentially subliminally) perceived norm. Thus, we are of the opinion that the data analysis in and of itself can be a valuable first step towards understanding profile construction across social networks. Disambiguating between users' conscious ideas about fitting in and actions observed through a data-based approach would require careful research design and can be the subject of a follow-on work.

References

An, J.; Kwak, H.; and Jansen, B. J. 2017. Automatic generation of personas using youtube social media data. In *HICSS*.

Ashforth, B. E.; Kreiner, G. E.; and Fugate, M. 2000. All in a day's work: Boundaries and micro role transitions. *Academy of Management review* 25(3):472–491.

Bakhshi, S.; Shamma, D. A.; and Gilbert, E. 2014. Faces engage us: Photos with faces attract more likes and comments on instagram. In *ACM CHI*, 965–974.

boyd, d. 2011. Social network sites as networked publics: Affordances, dynamics, and implications. In *Networked Self: Identity, Community, and Culture on Social Network Sites*. Routledg.

Chawla, S. K.; Khan, Z. U.; and Cornell, D. W. 1992. The impact of gender and dress on choice of cpa's. *Journal of Applied Business Research* 8(4):25.

Cimino, M. 2009. *Netiquette (on-line Etiquette): Tips for Adults & Teens: Facebook, MySpace, Twitter! Terminology–and More*. PublishAmerica.

Clark, S. C. 2000. Work/family border theory: A new theory of work/family balance. *Human relations* 53(6):747–770.

Coats, E. J., and Feldman, R. S. 1996. Gender differences in nonverbal correlates of social status. *Personality and Social Psychology Bulletin* 22(10):1014–1022.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE CVPR*, 248–255.

Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; and Darrell, T. 2013. DeCAF: A deep convolutional activation feature for generic visual recognition.

Edwards, K. 1987. Effects of sex and glasses on attitudes toward intelligence and attractiveness. *Psychological Reports*.

Farnham, S. D., and Churchill, E. F. 2011. Faceted identity, faceted lives: social and technical issues with being yourself online. In *CSCW*, 359–368. ACM.

Fölster, M.; Hess, U.; and Werheid, K. 2014. Facial age affects emotional expression decoding. *Frontiers in Psychology*.

Goffman, E. 1959. The presentation of self in everyday life.

Greenwood, S.; Perrin, A.; and Duggan, M. 2016. Social media update 2016. *Pew Research Center*.

Gross, J. J.; Carstensen, L. L.; Pasupathi, M.; Tsai, J.; Götestam Skorpen, C.; and Hsu, A. Y. 1997. Emotion and aging: experience, expression, and control. *Psychology and aging* 12(4):590.

Hogg, M. A. 2006. Social identity theory. *Contemporary social psychological theories* 13:111–1369.

International Markets Bureau, Canada. 2012. Global consumer trends - age demographics. Available at http://publications.gc.ca/pub?id=9.574883&sl=0, last accessed 8 March 2017.

Jain, V., and Learned-Miller, E. G. 2010. Fddb: A benchmark for face detection in unconstrained settings. *UMass Amherst Technical Report*.

Jang, J. Y.; Han, K.; Shih, P. C.; and Lee, D. 2015. Generation like: Comparative characteristics in instagram. In *CHI*, 4039–4042. ACM.

Jenkins, R., and Burton, A. 2008. 100% accuracy in automatic face recognition. *Science* 319(5862):435–435.

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

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*, 1097–1105.

Lim, B. H.; Lu, D.; Chen, T.; and Kan, M.-Y. 2015. # mytweet via instagram: Exploring user behaviour across multiple social networks. *ASONAM*.

Liu, H. 2007. Social network profiles as taste performances. *Journal of Computer-Mediated Communication* 13(1):252–275.

McCall, G., and Simmons, J. 1978. *Identities and interactions: An examination of associations in everyday life (revised ed.)*. Free Press.

McClure, E. B. 2000. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological bulletin* 126(3):424.

McLaughlin, C., and Vitak, J. 2012. Norm evolution and violation on facebook. *New Media & Society* 14(2):299–315.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Mod, G. B. B. 2010. Reading romance: The impact facebook rituals can have on a romantic relationship. *Journal of comparative research in anthropology and sociology* (2):61–77.

National Eye Institute, U.S. 2010. Statistics and data: Myopia. Available at https://nei.nih.gov/eyedata/myopia, last accessed 8 March 2017.

National Eye Institute, U. 2010. Facts about myopia. Available at https://nei.nih.gov/health/errors/myopia, last accessed 8 March 2017.

Ozenc, F. K., and Farnham, S. D. 2011. Life modes in social media. In *CHI*, 561–570. ACM.

Roberts, L. M. 2005. Changing faces: Professional image construction in diverse organizational settings. *Academy of management review* 30(4):685–711.

Rui, J., and Stefanone, M. A. 2013. Strategic self-presentation online: A cross-cultural study. *Computers in Human Behavior* 29(1):110–118.

Sebastian, R. J., and Bristow, D. 2008. Formal or informal? the impact of style of dress and forms of address on business students' perceptions of professors. *Journal of education for business* 83(4):196–201.

Sharda, N. 2009. *Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design: Visual Travel Recommender Systems, Social Communities, and User Interface Design*. IGI Global.

Skeels, M. M., and Grudin, J. 2009. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In *GROUP*, 95–104. ACM.

Souza, F.; de Las Casas, D.; Flores, V.; Youn, S.; Cha, M.; Quercia, D.; and Almeida, V. 2015. Dawn of the selfie era: The whos, wheres, and hows of selfies on instagram. In *ACM COSN*.

Strano, M. M. 2008. User descriptions and interpretations of self-presentation through facebook profile images. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 2(2):5.

Tifferet, S., and Vilnai-Yavetz, I. 2014. Gender differences in facebook self-presentation: An international randomized study. *Computers in Human Behavior* 35:388–399.

Tufekci, Z. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *AAAI ICWSM*.

Van Dijck, J. 2013. 'you have one identity': performing the self on facebook and linkedin. *Media, Culture & Society* 35(2):199–215.

Viola, P., and Jones, M. J. 2004. Robust real-time face detection. *Int'l. J. Comput. Vision* 57(2):137–154.

Walther, J. B.; Van Der Heide, B.; Kim, S.-Y.; Westerman, D.; and Tong, S. T. 2008. The role of friends appearance and behavior on evaluations of individuals on facebook: Are we known by the company we keep? *Human Communication Research* 34(1):28–49.

Wright, J.; Yang, A.; Ganesh, A.; Sastry, S.; and Ma, Y. 2009. Robust face recognition via sparse representation. *IEEE TPAMI* 31(2):210–227.

Zhao, S.; Grasmuck, S.; and Martin, J. 2008. Identity construction on facebook: Digital empowerment in anchored relationships. *Computers in human behavior* 24(5):1816–1836.

Zhao, X.; Lampe, C.; and Ellison, N. B. 2016. The social media ecology: User perceptions, strategies and challenges. In *CHI*, 89–100. New York, NY, USA: ACM.

Zhong, C.; Salehi, M.; Shah, S.; Cobzarenco, M.; Sastry, N.; and Cha, M. 2014. Social bootstrapping: how pinterest and last. fm social communities benefit by borrowing links from facebook. In *WWW*, 305–314.

Zhou, E.; Fan, H.; Cao, Z.; Jiang, Y.; and Yin, Q. 2013. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In *IEEE ICCV Workshops*, 386–391.