

Generation *Like*: Comparative Characteristics in Instagram

Jin Yea Jang, Kyungsik Han, Patrick C. Shih, and Dongwon Lee
College of Information Sciences and Technology
The Pennsylvania State University
{jzj157, kuh178, patshih, dongwon}@ist.psu.edu

ABSTRACT

The emergence of social media has had a significant impact on how people communicate and socialize. Teens use social media to make and maintain social connections with friends and build their reputation. However, the way of analyzing the characteristics of teens in social media has mostly relied on ethnographic accounts or quantitative analyses with small datasets. This paper shows the possibility of detecting age information in user profiles by using a combination of textual and facial recognition methods and presents a comparative study of 27K teens and adults in Instagram. Our analysis highlights that (1) teens tend to post fewer *photos* but highly engage in adding more *tags* to their own photos and receiving more *Likes* and *comments* about their photos from others, and (2) to post more *selfies* and express themselves more than adults, showing a higher sense of self-representation. We demonstrate the application of our novel method that shows clear trends of age differences as well as substantiates previous insights in social media.

Author Keywords

Teens; age difference; social media; Instagram

INTRODUCTION

Social media allows people to access, create, and interact with a wide range of information. Although this is true for people at all ages, teens' engagement in social media is especially high. Research has been conducted to understand teens in social media mainly in two ways. First, reports of ethnographic inquiries (e.g., interviews, observations, focus groups, etc.) found that teens spend a considerable amount of time interacting and socializing in online spaces, and use social media as a main channel to share their everyday life, learn something new and useful, and establish and reinforce social empowerment [5,7]. Teens also shift to new social platforms quickly to meet their fast changing priorities, and personal and social needs [1,5]. The second method utilizes quantitative analysis of real usage data. For example, in an analysis of a total of 100 users, [11] showed that teens express themselves and share personal stories online more

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04...\$15.00
<http://dx.doi.org/10.1145/2702123.2702555>

than senior adults (over 60). They reported that teens use diverse social media sites and maintain multiple forms of communications, because each channel affords different ways of interacting with others. [13] analyzed interactions by a total of 160 users in a public online news community and found that teens maintain more social connections than the elderly (over 65).

Although these two approaches show many perspectives on how teens use and perceive social media compared to other age groups, very few empirical studies have been conducted using large datasets. Many studies are based on relatively small sample size (e.g., 5-30 users for qualitative studies and 100-300 for quantitative studies). As people create a huge volume of social media data everyday, we believe that there are novel opportunities to explore trends with large datasets. However, thus far, this attempt has found to be challenging, because most social media sites neither collect nor reveal users' age information.

In this paper, we introduce a novel method for identifying age information from user profiles. Recent research has shown the possibility of using a facial detection technique to extract the age information from users' photos [3]. We enhance this method using natural language processing of users' bio statements, which allowed us to collect a total of 27K teens and adults in Instagram. With this data, this paper aims at exploring the following question.

How do teens use and engage in Instagram compared to adults?

Our contributions include: (1) demonstrating the possibility of leveraging a combination of textual and facial recognition methods to accurately detect users' age information, (2) extracting a large dataset for conducting comparative analysis of social media usage based on age differences, (3) identifying clear trends of age differences in social media that are otherwise unattainable using the qualitative and quantitative methods employed in previous studies. We close this paper by proposing future research opportunities afforded by our method.

STUDY DATA COLLECTION

We chose Instagram, because over 90% of Instagram users are under the age of 35, which is suitable for our target age groups of teens and adults¹. We used the programming API²

¹ <http://www.businessinsider.com/instagram-demographics-2013-12>

² <http://instagram.com/developer/>

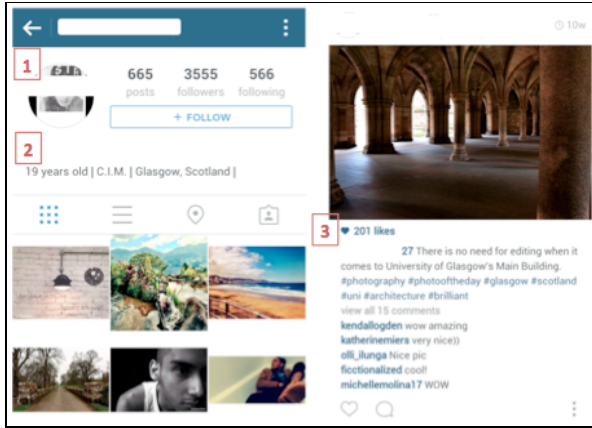


Figure 1. Screenshot of the user profile (anonymized) in Instagram, including (1) a profile image, # posts, followers, and followings; (2) bio; (3) # Likes, tags, and comments.

to extract users' all usage data. The data collection was done between April and May 2014. We first chose one random seed user and crawled followers of the seed user until we have 150K users. We then randomly chose 1K users from the pool of 150K and again crawled followers of them until we reached 2 millions of unique users. We used this two-step random-seed crawling process to minimize the bias in sampling a homogenous population. Overall, the dataset includes various pieces of user information such as name, the number of photos posted, the number of Likes, tags and comments in the photo, the number of followers and followings, and a bio description (Figure 1).

We define our target user populations as follows:

- *Teens*: people who are between 13 and 19
- *Adults*: people who are between 25 and 39

Classifying users to a specific age group was challenging, because most social media platforms including Instagram neither collect nor disclose users' age information. We propose a novel method that leverages two existing media contents (i.e., bio descriptions and profile images) with existing APIs. First, we applied textual pattern recognition algorithms to parse a list of patterns that specifically describe users' age in the bio (e.g., "I am a teenager," "I am 17 years old," "I'm 23"). Second, we used an online tool called Face++³, which was designed to detect and provide the age and gender information of people presented in the photo. Face++ provides an API and has been utilized to extract age information in recent studies, and age verification showed a high accuracy (i.e., 96% accuracy in age range between 18 and 35 and 99% in ages over 35) [3].

Based on the users' bio descriptions, we identified 13,533 teens and 8,596 adults. To fill the gap between two groups, we analyzed the rest with Face++ and added 4,756 more

adults, resulting in a total of 26,885 teens and adults for the analysis. Finally, we manually verified the age of all users to make sure the data accurately represented each group. A total of five people (two authors and three Turks) visited all users' Instagram page and checked their profile image, bios and posted photos. We only included users who achieved agreements among more than four raters for the analysis. Note that Face++ was only used for detecting adults but not teens in this study, because facial recognition was less accurate for people of a younger age range.

STUDY RESULTS

Difference in engagement

	Teens (13,533)		Adults (13,352)	
	Median	SD	Median	SD
# Photos	110	272	175	487
# Likes	3,293	29,851	2,150	24,829
# Tags	446	2,595	294	2,511
# Comments	175	1,016	35	1,023
# Followers	401	3,683	348	5,700
# Followings	286	2,045	272	2,699

Table 1. Summary of interactions & activities by two groups.

Table 1 shows the summary of Instagram activities for each group. Because all variables show a long-tailed distribution, we used the median value for the analysis. Because of our large and non-normalized datasets, to correctly measure the differences, we used eta-square (the effect size), denoted as η^2 , which refers to the proportion of variance associated with each of the main effects, interactions, and error in an ANOVA study [12]. As a rule of thumb, $\eta^2 = 0.01, 0.13, 0.26$ are considered to be small, medium and large respectively. The results indicate all significant differences where teens post fewer photos ($\eta^2 = 0.02, p < 0.001$) but have more Likes ($\eta^2 = 0.04, p < 0.001$), tags ($\eta^2 = 0.01, p < 0.05$), and comments ($\eta^2 = 0.02, p < 0.001$) than adults.

	Teens (13,533)	Adults (13,352)
# Likes / # Photos	56.10	40.03
# Tags / # Photos	6.34	4.70
# Comments / # Photos	2.52	1.06

Table 2. The ratio of Likes, tags, and comments to photos.

We also measured the ratio of Likes, Tags, and Comments to Photos (Table 2). As a result, teens are likely to receive more Likes per photo than adults ($\eta^2 = 0.09, p < 0.001$). Teens add more tags ($\eta^2 = 0.01, p < 0.001$) per photo than adults. One of tagging motivations is to attract more views [2], and this implies that teens may want to make their photos and themselves more exposed to others than adults. Teens also have more comments ($\eta^2 = 0.07, p < 0.001$) per photo than adults, showing higher interactions in Instagram.

The result of teens having fewer photos than adults is quite unexpected and interesting; thus, we conducted additional analyses based on our sample. First, we hypothesized that teens may have limited resources to explore environments outside their daily activities. To test this, we identified latent topics from the tags of users' photos through an LDA analysis [4] using Mallet [9]. We used tags to infer photo

³ <http://www.faceplusplus.com/>

content, because people tend to add tags that represent the photos they post [6]. We also obtained ground-truth tag topics from two popular websites (i.e., tagsforlikes.com and tagstagram.com) and then manually coded the types of photo topics from Mallet’s output into those topics. Table 3 summarizes the 11 topics extracted from our dataset.

Topic	Tag examples
Arts/photos/design	photo, interior, architect, design, building
Entertainment	music, movie, pop, rock, song, star, dance
Fashion/beauty	makeup, model, fashion, jewelry, beauty
Follow/like	followme, followback, follow, tagsforlike
Foods	food, coffee, yummy, delicious, dessert
Instagram-tags	instagood, instalove, instadaily, instashare
Locations	nyc, boston, spain, brazil, dutch, europe
Mood/emotion	love, happy, depressed, bored, sad, great
Nature	sky, sun, ocean, beach, flower, sunset
Social/ people	family, girlfriend, friends, folks, gay, pets
Sports/wellness	hiking, biking, fitness, cleaneating, soccer

Table 3. LDA-discovered topics from all users (N=26,885).

We then calculated the ratio of topics that users in each group have in their photos, as presented in Table 4.

Topic	Teens (13,533)	Adults (13,352)
Arts/photos/design	4.59%	16.99%
Entertainment	11.34%	6.28%
Fashion/beauty	1.53%	4.20%
Follow/like	11.43%	2.22%
Foods	3.18%	3.13%
Instagram-tags	4.10%	3.22%
Locations	7.05%	24.89%
Mood/emotion	40.54%	14.21%
Nature	5.05%	9.04%
Social/people	4.50%	10.74%
Sports/wellness	6.69%	5.08%

Table 4. LDA-discovered topics from all teens and adults.

We found a clear difference between two groups in terms of topic types. On the one hand, for teens, more than half of photos are in the “Mood/emotion” and the “Follow/Like” topics. These topics are not necessarily tied to the content of photos but rather describe one’s emotional status or intention to have more followers. On the other hand, adults show a high ratio in more diverse topics, including “Arts/photos/design,” “Locations,” “Mood/emotion,” “Nature,” and “Social/people.” These topics imply more diverse contents in the photos, including photos that depict different facets of cities and countries around the world, photos of arts and design (where some of them were taken professionally), photos of a variety of people, and so on. This may be due to the fact that teens are financially and culturally dependent on their parents to venture outside of their daily activities compared to adults.

Second, previous studies found that many teens manipulate the content that they have shared; for example, removing photos [1,8]. To check this, we additionally collected the number of photos by teens and adults (13K each) in the 12-hour interval over 12 days (from December 26, 2014 to January 6, 2015). We then calculated the delta of photo counts in every two consecutive time slots and checked the number of users who removed their photos in 12 hours. As

a result, given the same number of users for each group, more teens tend to remove photos than adults ($t(44) = 41.9$, $p < 0.001$; Figure 2), supporting our hypothesis.

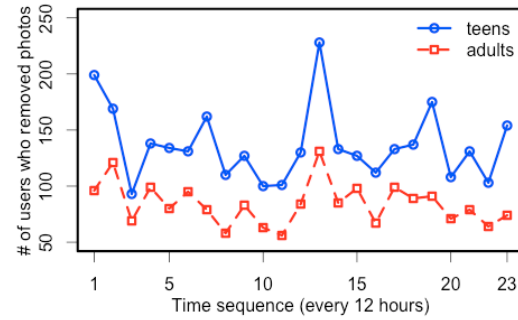


Figure 2. Number of users who removed photos (teens: 13K, adults: 13K).

Difference in self-representation

We extended our finding about teens’ high engagement by analyzing their photos. We assumed that teens might post photos of themselves, namely “selfies,” more than adults, because teens see social media as a place for self-representation [5,7]. To infer the photo content, we calculated the number of photos with tags and those with “selfie-tags” (we used #selfie and #me) as well as their ratio for each group. As a result, teens tend to post selfies more than adults ($\eta^2 = 0.08$, $p < 0.001$; Table 5).

	Teens (13,533)	Adults (13,352)
# Selfies / # Tagged-Photos	226,382 (12.8%)	99,782 (7.5%)

Table 5. Ratio of selfies for each group based on the tag analysis.

Our second approach to find out selfies was a photo content analysis. We verified photos posted by users in each group and calculated the percentage of selfies. With the randomly selected 1K users, we extracted around 56K photos from teens and 86K photos from adults. We manually checked all photos if they are selfies. On average, we found that teens have more selfies than adults ($\eta^2 = 0.05$, $p < 0.001$; Table 6). These results support the idea that self-representation is one of the motivations for teens to engage in social media over adults.

	Teens (1,000)	Adults (1,000)
Total # Photos	56,458	86,227
Total # Selfies	14,627 (25.9%)	14,823 (17.1%)

Table 6. Ratio of selfies from photos by randomly selected users based on the photo content analysis.

Lastly, we utilized the Linguistic Inquiry and Word Count (LIWC) [10] tool to parse words representing emotions in the bio descriptions. We randomly chose 8,596 teens and adults for the comparison of word per sentence (a level of expression) and words for follows (the number of words of “follow,” “following,” and “follows”).

Table 7 shows the results with two insights. First, teens’ bio contained more *word per sentence*, than adults ($\eta^2 = 0.06$, $p < 0.001$), indicating teens tend to be more verbose on their bio, which is a way of self-expression. Second, confirming

the findings in Table 4 that teens add more photos in the *Follow/Like* topic, teens tend to use more words, “follow, following, or follows” in their bio descriptions than adults ($\eta^2 = 0.03$, $p < 0.001$). A qualitative analysis of teens’ bio descriptions revealed that teens actively advertise others to follow them, which would also lead to gaining popularity, making friends, and making their photos more visible. For example, many teens post texts, similar to the following in their bios: “*Please follow me and I follow you.*”

	Teens (8,596)	Adults (8,596)
Word per sentence	16.11	10.62
Words for follow-words	2.39	1.44

Table 7. Scores of two attributes from the bio description. The higher result means more expressive and more words for follow-ships.

DISCUSSION AND CONCLUSION

The primary contribution of our work is the demonstration of a method that leverages existing tools (i.e., Instagram and Face++ APIs) and social media contents as a channel and triangulates findings about different age groups that can be difficult to observe using existing methods.

Our analysis shows an interesting trend that teens post fewer photos than adults. To explore reasons behind it, we further analyzed our dataset in two ways. First, the LDA-based topic analysis showed that over half of teens’ photos were in “Mood/emotion” (40%) and “Follow/Like” (11%), while adults’ photos were more evenly distributed across other topics. This result shows that teens tend to post topics related to themselves and focus on self-expression. Second, we found that more teens tend to remove photos than adults everyday, which substantiates some insights presented in previous studies. Many teens tend to manipulate their photo content to receive as many Likes as possible or sometimes remove some photos with too few Likes [1,8]. This may be because attention generated by Likes has become one way of establishing self-validation and self-worth⁴, and teens want to show off their “coolness” to the public [5].

We also found that teens tend to have more Likes, tags, and comments and be more expressive about themselves in their profiles and photos. This shows that teens are leveraging social media as a “conversation space,” and using Likes, tags, and comments to have conversations and interactions. This may be because teens are well aware of the intention of these activities as social signals and familiar with the technology and the “tagging culture” in online space, which reinforces teens’ social practices [1].

Although our study shows many insights, we acknowledge some limitations that can be handled in future studies. First, errors may exist in the detection of age information even if we manually verified them. Many users provide additional social media links (e.g., Facebook, Twitter, etc.) in their profile. Future studies applying our method should obtain additional age information from those sites. Second, our

results may only be limited to users in Instagram. We plan to extend our study to other social media sites to validate our method and compare results. Third, the age information from users’ profiles could be obsolete when social media becomes more mature. A possible way to address this is to validate users’ age information by comparing users’ most recent selfies to their profile photos. This could potentially address the information obsolescence issue, but a future study to validate its accuracy will be necessary.

Our findings also ask for follow-up qualitative studies with some detailed questions that will give additional insights on teens in social media; for example, “Why do teens post fewer photos?,” “What makes teens engage in Liking, tagging, and commenting?,” “To what extent do teens and adults consider social media a conversation space?,” etc.

Overall, we believe that our novel method opens up a new possibility of mining activities among different age groups in social media. We hope that this method can be used as a guide for researchers to better understand social phenomena and to formulate future research directions.

ACKNOWLEDGEMENTS

This research was in part supported by NSF CNS-1422215 award.

REFERENCES

1. Agosto, D. E., Abbas, J., & Naughton, R. Relationships and social rules: Teens’ social network and other ICT selection practices. *JASIST* 63, 6 (2012), 1108-1124.
2. Ames, M., & Naaman, M. Why we tag: motivations for annotation in mobile and online media. *CHI 2007*, ACM Press (2007), 971-980.
3. Bakhshi, S., Shamma, D. A., & Gilbert, E. Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. *CHI 2014*, ACM Press (2014), 965-974.
4. Blei, D. M., Ng, A. Y., & Jordan, M. I. Latent dirichlet allocation. *JMLR* 3, (2003), 993-1022.
5. boyd, d. Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. *MacArther Foundation Series on Digital Learning - Youth, Identity, and Digital Media*, (2008), 119-142.
6. Hollenstein, L., & Purves, R.S. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science* 1, 1 (2010), 21-48.
7. Ito, M., Baumer, S., Bittanti, M., et al. *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media*. MIT Press, 2009.
8. Madden et al. Teens, Social Media, and Privacy. *Pew Research Center*, 2013.
9. McCallum, A. K. “*Mallet: A Machine Learning for Language Toolkit*,” 2002.
10. Pennebaker, J. W. & Francis, M. E. *Linguistic inquiry and word count (LIWC)*. Mahwah, NJ: Erlbaum, 1999.
11. Pfeil, U., Arjan, R., & Zaphiris, P. Age differences in online social networking - A study of user profiles and the social capital divide among teenagers and older users in MySpace. *Journal of Computers in Human Behavior*, 25 (2009), 634-654.
12. Tabachnick, B. G., & Fidell, L. S. *Using Multivariate Statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon, 2001.
13. Zaphiris, P., & Sarwar, R. Trends, similarities, and differences in the usage of teen and senior public online newsgroups. *TOCHI* 13, 3 (2006), 403-422.

⁴ <http://www.pbs.org/wgbh/pages/frontline/generation-like/>