The Pennsylvania State University

The Graduate School

College of Information Sciences and Technology

TEENS IN SOCIAL MEDIA: DATA-DRIVEN COMPARATIVE ANALYSIS ON BEHAVIORS IN INSTAGRAM

A Thesis in

Information Sciences and Technology

by

Jin Yea Jang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2016

The Thesis of Jin Yea Jang was reviewed and approved* by the following:

Dongwon Lee
Associate Professor of Information Sciences and Technology
Thesis Advisor

Mary Beth Rosson
Professor of Information Sciences and Technology

Xiaolong (Luke) Zhang
Associate Professor of Information Sciences and Technology

Andrea H. Tapia
Associate Professor of Information Sciences and Technology
Graduate Program Chair

*Signatures are on file in the Graduate School

ABSTRACT

Social media plays an important role in connecting people and making them interact with each other. Its influence is more direct and stronger to teens, because they are more active and engaged in using social media than other generations. Much research has presented teens' high use and engagement as well as characteristics in social media; however, the way of analyzing them has mostly relied on ethnographic accounts or quantitative analyses with small datasets. In this thesis, I present the analysis of how teens use social media differently using a large dataset of Instagram. First, I present the characteristics of Instagram use by people with respect to structure, influence, and context, in order to see its general use patterns. Second, based on the methods used in the first study, I present how teens use and engage in Instagram. I describe a novel method to detect teens and adults using user profiles, mixing with textual and facial recognition approaches. With around 27,000 teens and adults identified, I present the comparative study of the two user groups. The study results highlight that (1) teens tend to post fewer but remove more photos, influenced by the number of Likes; (2) teens tend to be more engaged in Liking and commenting and to express their emotions and social interests, than adults; (3) teens tend to have less diverse photo content. Lastly, I will discuss theoretical and practical interpretations and implications of the results, as well as future directions of my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Social media has been widely adopted in people's daily lives, especially through the help of mobile devices, allowing them to access, create, and interact with a wide range of information. In particular, teens are known to be highly engaged in social media (Adams & Marshall, 1996; Grinter, Palen, & Eldridge, 2006). It has been reported that 47 percent of all American teens now use a smartphone, 81 percent of teens use social media, and 93 percent of teens are online (Duggan & Brenner, 2013; Madden et al., 2013). Most noteworthy is the phenomenon that teens and young adults appear to be early adopters — and arguably the most active users — of social media (boyd, 2008). For them, social media has become the new channel and new way of representing themselves (Ong et al., 2011), to share their everyday activities and thoughts with friends (Ellison, Steinfield, & Lampe, 2007), to establish and maintain social connections (Muscanell & Guadagno, 2011), and to learn something new and useful (Ito et al., 2008).

Teens are still believed to be the most engaged and adventurous among all users in social media. On the one hand, being acclaimed as the "digital natives" (Prensky, 2001), teens grow up with an abundance of communication technology and therefore are believed to be more technologically-savvy than adults. On the other hand, from a developmental perspective, teens may consider social media as an exciting opportunity for social interaction space (Ito et al., 2008) and self-display (Livingstone, 2008), while adults may be more concerned about their information privacy in online disclosure. Moreover, given that socialization is an especially influential process in childhood and adolescence, interaction with their peers through social media plays an important role in teens' life and has a significant impact on teens' self-esteem and psychological well-being (Valkenburg, Peter, & Schouten, 2006). Teens' social needs drive their social media behaviors — they would therefore stay active online in order to build and maintain connections with their peers through online interactions.

However, both the assumption of teens being active users in social media and the rationale behind this assumption have not been sufficiently studied and validated through their real use of social media. First, despite the growing body of work that examines teens' online behaviors and technology use, little effort has been put into directly comparing teens' and adults' social media use and activities. Therefore, it is difficult to determine if teens' actual use of social media is unique compared to other age groups. Second, existing studies of social media use have been mostly limited to ethnographic accounts (e.g., interviews, focus groups, etc.) or self-reported survey studies, while empirical investigations of large-scale user data is lacking. The latter is particularly useful for developing an understanding of the unique behavioral patterns of teen social media users and the underlying strategies that they may use to manage their online expressions. However, such an approach faces technical challenges. For example, identifying the teen versus adult users in social media is non-trivial, because many users often do not publicly reveal their age information nor, in many cases, do social network sites (SNS) ask for age information at the time of registration.

In this regard, I seek to address the aforementioned limitations of social media studies and investigations. I first explore and identify general aspects of social media use patterns from a large number of users. Then I switch my focus to teen populations generating research hypotheses and analyzing behavioral patterns based on the insights obtained from the first study. The outline of this thesis is as Figure 1-1.

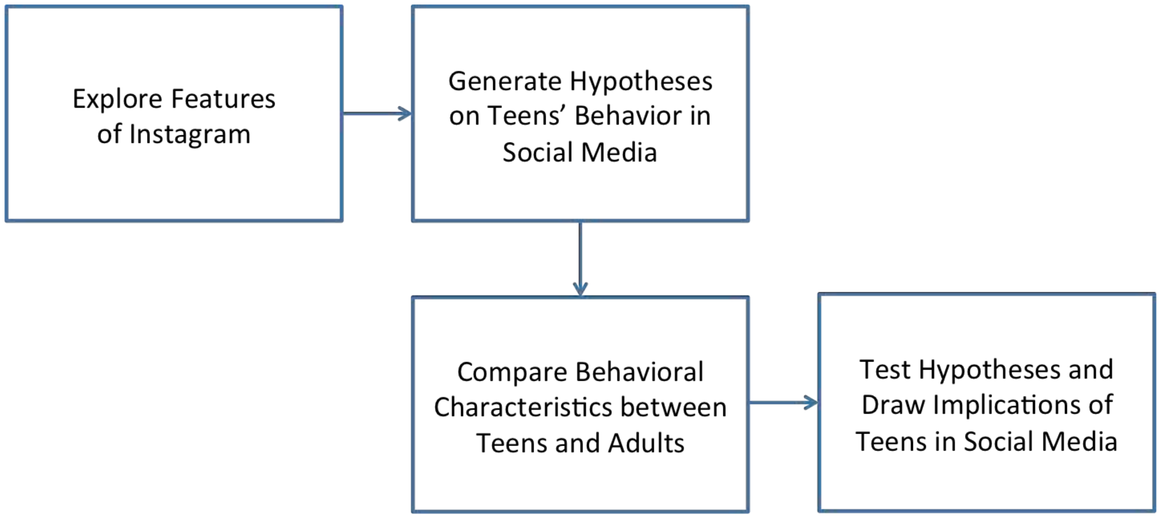


Figure 1-1: The workflow of the study.

First, I investigate the characteristics of Instagram to see the general usage patterns with the datasets of 20 million users and their activities. I employ three main research questions with respect to structural, influential, and contextual aspects as follows: What are the structural characteristics of user activities? (Structure); How do the activities (photos, Likes, tags, comments, followers, and follows) influence each other? (Influence); What are the contextual characteristics of the activities? (Context)

Second, I introduce a new hybrid method of textual pattern matching and facial recognition to detect users' age information in a large scale. Recent research has shown the possibility of using a facial detection technique to extract the age information from photos (Bakhshi et al., 2014). I enhanced this age detect method using natural language processing of users' bio statements. With this method, I collected user information and usage data from a total of 27,000 teens and adults in Instagram.

Lastly, I present some preliminary comparative analyses between teens and adults. I found that teens tended to have fewer photos than adults because of limited topics and photo removal. I also found that teens tended to have high engagement in Instagram.

Overall, the followings are the main contributions of my work in this thesis.

- First, I outline two hypotheses based on the developmental literature and related work, serving as a theoretical foundation for explaining how age factors in social media behaviors.
- Second, this study further uses the temporally extended, large-scale dataset in order to empirically analyze the way teens and adults use Instagram over time, and the different patterns of interacting with other users that teens and adults show.
- Third, based on the empirical findings, I draw theoretical and practical interpretations and implications that may provide useful insights and guidance for future research and design.

Compared to existing research in this domain, my work is among the first to conduct an analysis with large-scale user activity data to extensively reveal behavioral patterns and empirical understandings as well as to identify potential factors that drive these patterns.

Chapter 2

Insights from previous social media studies

First, much research has presented various structural aspects of one's online social network, ranging from its component and formation to its comparison to other network types. For example, (Gruzd et al., 2011) showed that, in social media, people form interlinked personal communities based on their follow and following connections as well as the norms, languages, and techniques used by them within the network. Somewhat differently, (Chen et al., 2014) argued that not all members are fully connected with each other and many relationships are missing in online social networks. They presented a new structure-based approach that leverages social communications (i.e., posts and replies) among users to identify different communities in which they engage. (Huberman et al., 2009) found that a followship link between any two people in social media was not positively related to a network of people whom they actually interact with. They emphasized the importance of eliciting a hidden social network that goes beyond simple follow-based relationships. By taking a similar approach, I will show how a Like network formed from Like activities is structured and developed, which has not been articulated in social media research.

Second, when it comes to the influence of online social media, I was in particular interested in the extent to which different elements that exist in a social media platform influence one particular element in the same platform. There are a number of studies that detail those relationships. For example, (Lampe et al., 2007) explored the different levels of influence of profile elements on the number of friends on Facebook. They found that the number of friends was positively associated with several common referents, such as high school, hometown, same major, and same school, even after controlling for gender, time on the system, and the updated time.

Similarly, in the study on Pinterest, (Gilbert et al., 2013) posited that being female, having fewer followers, and using four specific verbs (i.e., use, look, want, and need) will lead to having more re-pins. In Twitter, studies have found that having tags and URLs show the strongest effects on having more retweets (Suh et al., 2010). Similar to these studies, I also aimed at exploring the relationship between the number of Likes and other elements that specifically pertains to "Instagram design interfaces," including the number of photos, comments, tags, followers, and followings. These are the direct indicators of one's engagement and activities in Instagram.

Lastly, much research has investigated the contextual aspect of social media. For instance, it has been found that social media creates a communication space for presidential elections (Tumasjan et al., 2010), workspaces (Davison et al., 2014), and major incidents or disasters (Vieweg et al., 2010) (Yardi & boyd, 2010). Studies have also indicated that social media reengineers the way of interactions between doctors and patients (Hawn, 2009), provides richer local information to residents and facilitates local interactions (Schroeter, 2012), and helps teachers maintain professional ties with different educational communities as well as share resources and make connections with students (Moran et al., 2011). Based on these studies, I found that, in most cases, the contextual information was obtained from the text-based content. However, Instagram is different, because it is a photo-based social media platform. (Hu, 2014) presented content categories from Instagram photos; however, the small sample size (200 photos from 50 users) used in the analysis limits their findings. In my study, to infer its content, I decided to leverage tags, because previous research has reported that users tend to add tags that meaningfully describe the photo content (Hollenstein & Purves, 2010). With this rationale, I have applied a probabilistic topic model-based tag analysis and measured the relationship between photo topics and Likes.

Extracting photo information also allowed me to study an additional contextual aspect. One study method utilized in many social media studies is to articulate different use cases by different groups. For example, (Ottoni et al., 2013) analyzed gender roles and behaviors in Pinterest, and found that females tend to have more diverse interests but males tend to be more interested in specific topics. (Haferkamp et al., 2012) studied self-presentation in social media, and found that females are more likely to use online social networking sites for comparing themselves with others, while males tend to use them to find friends.

Interestingly, little research has presented the characteristics of specific age groups (e.g., teens) in social media, mainly because of a limitation on identifying user's age information with most existing technologies. By using a hybrid method to identify teens and collect their usage logs (which will be described in the later section), I extended the insights obtained from my first Instagram study to understand how teens use and engage in Instagram.

Chapter 3

Characteristics of Instagram

Instagram was chosen for this study for two main reasons. First, Instagram is one of the most popular SNS with users who create and share mainly photos everyday. Because of its high popularity, there has been a great volume of research studies on Instagram. Examples include exploring the relationship between photo content and engagement (Bakhshi, Shamma, & Gilbert, 2014), analyzing photo content and user types (Hu, Manikonda, & Kambhampati, 2014). Second, given the fact that more than 90 percent of Instagram users are under the age of 35 (Duggan & Brenner, 2013), it is suitable to study my target age groups of teens and adults.

Prior to see the generation behavioral differences, the investigation of Instagram is preceded to take a look at what kinds of characteristics this social media service has. The results of this investigation can use as guidelines when comparing the two age groups.

The followings are the main questions in this chapter:

- *Structure:* What are the structural characteristics of user activities?
- *Influence:* How do the activities (photos, Likes, tags, comments, followers, and follows) influence each other?
- *Context:* What are the contextual characteristics of the activities?

Data collection

I used the programming API[1] to extract usage data for all users. The data collection was done between March and May 2014. I first chose one random seed user and crawled the followers of the seed user until I collected 150,000 users. I then randomly chose 1,000 users from the pool of

[1] http://instagram.com/developer/

150,000 users and again crawled the followers of 1,000 users until I reached 2 million unique users. I used this two-step, random-seed crawling process to minimize the bias in sampling a homogenous population.

To reduce biases from the data collection and speed up subsequent data analysis, then, I randomly generated several subsets of users of different sizes (e.g., 1K, 10K, 100K, and 500K users) from 20M users and used them for different analyses (see Figure 3-1). For instance, to answer the structure question, I used 500K users to understand overall usage activities, but used 100K users to closely monitor their daily usage over a month. In addition, I used 10K users to calculate the portion of Likes from one's followers and 1K users to generate and compare Like and Follow Networks at a more fine-grained level. For the contexts question, I used 100K random users.

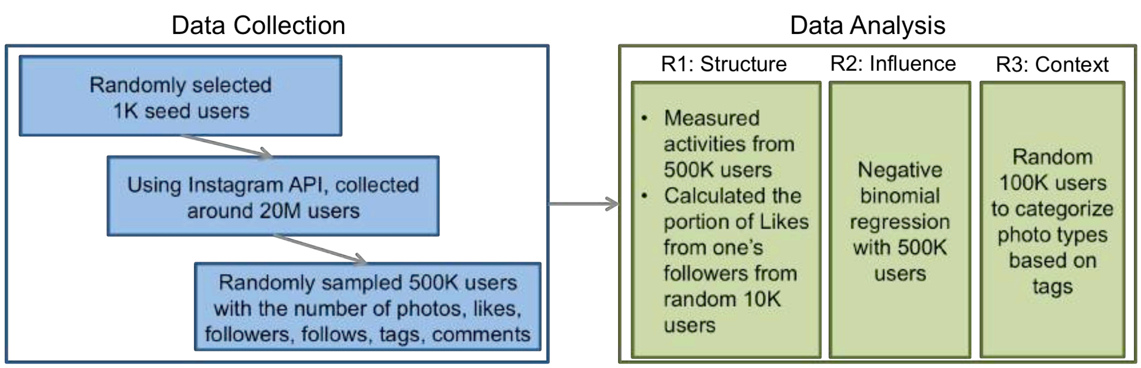


Figure 3-1: Workflow of data collection and analysis.

By using different sizes of datasets for different measurements, I was able to handle different data formats required for a particular study and speed up the processing time for analyzing data. Otherwise, for instance, processing 20 million users and 2 billion Likes in the base dataset was prohibitively time-consuming and highly resource-intensive. The datasets that I collected from Instagram consisted of seven types as follows:

- *Posters*: (Instagram) users who posted/uploaded photos

- *Photos*: Posters' photos
- *Likes*: Likes added to posters' photos
- *Tags*: Tags added to posters' photos
- *Comments*: Comments added to posters' photos
- *Followers*: Users who follow posters
- *Follows*: Users whom posters follow

Note that Photos, Tags, Followers, and Follows pertain to posters, whereas Likes and Comments are added from other users who access posters' photos.

Results

The results of the study are composed of three sections: structure, influence, and context.

Structure

Table 3-1: Descriptive statistics of the dataset (N=500K).

Variable	Median	Mean	Max	S.D.
# Photos	166	309	57,925	487
# Likes	1,984	11,122	61,606,804	224,292
# Tags	21	228	97,249	1,034
# Comments	58	320	1,112,862	3,861
# Followers	623	2,404	2,751,722	16,488
# Follows	292	734	5,291,779	19,026

Table 3-1 shows the basic statistics of six variables from 500K posters. In general, there is a wide spectrum of variances in all variables, indicated by their high standard deviations. In particular, the variance for the number of Likes is higher. While there are many users whose photos have received no Likes at all, for instance, there is a user whose photos have garnered as many as 61 million Likes. In general, the number of Likes that a user or a photo has received shows a typical long-tail distribution with only a small fraction of dominating users or photos. As empirical evidences, four graphs in Figure 3-2 illustrate different functional relationships between the number of Likes and: (1) the number of users, (2) user rank, (3) the number of photos, and (4) photo rank. Note that (1) and (2) are based on all 500K users, and (3) and (4) are based on around 5M photos posted by randomly selected 100K users. While not identical, all four graphs exhibit similar characteristics—i.e., power-law distributions hold approximately only over a limited range (as the number of Likes increases on x- axis), and only a small number of dominating users or photos receive a disproportionally large number of Likes.

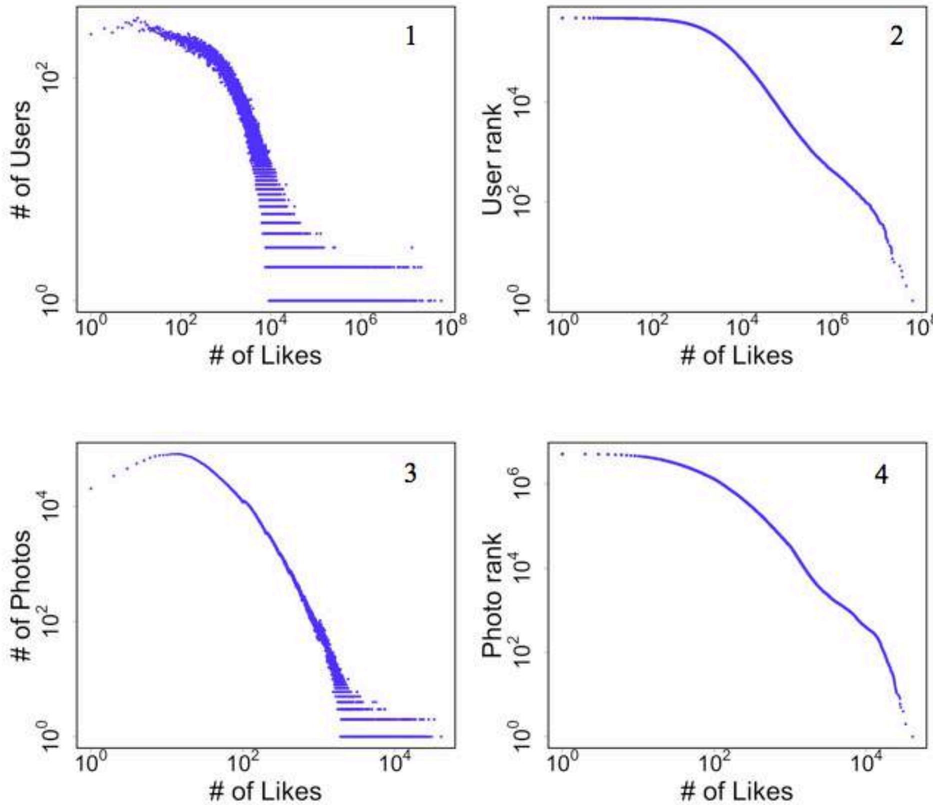


Figure 3-2: Distributions of # of Likes with respect to: (1) # of user, (2) user rank, (3) # of photos, and (4) photo rank. (1) and (2) are based on 500K users, and (3) and (4) are based on around 5M photos posted by randomly selected 100K users. All graphs in log-log plot.

Next, I investigated the number of Likes that a poster receives from other users. From randomly chosen 10K posters, I calculated the ratio of Likes added by each poster's followers. With the IDs of users who added Likes and those of followers, I was able to check the percentage of Likes received by one's followers. As a result, interestingly, I found that almost a half of Likes were from random users with no follow-relationship (Table 3-2). This in part indicates that Instagram users not only check photos from people they follow, but they also navigate random photos and simply add Likes if they like those photos.

Table 3-2: Total number of Likes from all users and the ratio of Likes from one's followers (N=10K).

	Mean	**Median**
# Photos per user	47	29
# Likes per user	1,333	1,009
# Likes from one's followers per user	742 (55.6%)	476 (47.1%)

Having a single Like could be explained by a unique characteristic of Instagram. Instagram users can easily access many random photos shared by random users. For example, Instagram provides photo pages that display hot photos (e.g., photos of this month), or photos by specific hashtags (e.g., #halloween, #christmas), making it easy for any user to add Likes to random photos or receive Likes from random users.

Influence

The second question examines how do the activities (photos, Likes, tags, comments, followers, and follows) influence each other. Among the activities, Liking is one of most important interaction behaviors in Instagram. Adding Likes shows users' interests in the content or the user who posted the content. It implies personal preference or interest in a media shared by others. The number of Likes simply shows a reputation of media.

I investigated the factors that influence the number of Likes. My assumption was that a poster might receive "more" Likes due to: (1) the posting of many interesting photos, (2) having many followers or follows, (3) the addition of many tags to photos (which are used in the "search" feature), or (4) the people's tendency to add Likes while adding comments to photos. Articulating the relationships among these factors may provide some perspectives of the activities. For this analysis, I used a negative binomial regression model, which is a statistical method to model Like activities by considering other variables as the predictors. This model has been previously used to understand the relationship among variables in other social networking sites (Bakhshi et al., 2014; Gilbert et al., 2013). The fact that the dependent variable, which is the number of Likes, is a count and conditional variance of each variable that exceeds its conditional mean suggests that using the negative binomial regression model is appropriate. I used STATA software for the analysis.

Table 3-3: The result of the negative binomial regression. The dependent variable is # of Likes, which is also countable (N: 500K; IRR: Incident Rate Ratio).

Variable	**β**	**IRR**	**Std. err.**	**z**	**p**
Followers	0.079	1.082	0.0004	173.0	< 0.0001
Photos	0.046	1.047	0.0004	102.3	< 0.0001
Comments	0.032	1.033	0.0002	114.5	< 0.0001
Tags	0.028	1.028	0.0002	120.7	< 0.0001
Follows	-0.005	0.994	0.0005	-9.7	< 0.0001

Note: Alpha (estimate of the dispersion parameter): 1.40, Likelihood-ratio test of chi-square: 2.3e+0.9, p < 0.0001.

Table 3-3 presents the result of the negative binomial regression where p-value indicates that the model is statistically significant ($p < 0.0001$) and the number of Likes is the dependent variable. The alpha value of the model refers to the estimate of the dispersion parameter, and the fact that alpha is greater than zero (1.40) indicates that the data are over dispersed and better estimated using a negative binomial model than a Poisson model. The model also shows the large test statistic of the likelihood-ratio chi-square test, again indicating that using the negative binomial model is appropriate.

The IRR (Incident Rate Ratio) result refers to the change in the dependent variable in terms of a percentage increase or decrease, which measures the effects of the independent variable on the dependent variable. More specifically, the IRR for followers (1.082) means that for each one-unit increase in followers, the expected number of Likes increases by 8.2% (p <

0.0001), while holding the other variables in the model constant. This in part indicates that people are likely to add Likes to the photos posted by those whom they are following, and having more followers is likely to lead to having more Likes.

Likewise, the expected number of Likes increases by 4.7%, 3.3%, and 2.8% with every one-unit increase in photos, comments, and tags, respectively ($p < 0.0001$), while holding the other variables in the model constant. For photos, although I had expected to see a higher percentage of its influence on the number of Likes (i.e., more photos, more chances to get Likes), the results still show a relatively significant effect. Commenting is another (could be more explicit) way of expressing one's thought, and people might add Likes while adding comments. In addition, the result for tags seems to be supported by an interesting culture in Instagram where the tags can be used as a way of promoting oneself or one's photos, similar to the way hashtag (#) is used in Twitter. Lastly, regarding follows, it shows a negative effect (-0.5%) on the Like count. This result can be partly explained by the fact that many popular and active posters that have many followers (i.e., they also tend to receive many Likes) do not always follow back with a similar number of others. This also implies that following more people does not always guarantee receiving more Likes back from those people.

In summary, the results show that all independent variables, except follows, are positively related to the number of Likes to different extents. Especially, I found that having more followers and adding more photos seem to be more influential with respect to having more Likes.

Context

The third question explores the contextual aspects. In particular, I extract contextual information from photos by means of the tags in photos. Topic models are often useful for analyzing a large collection of unlabeled texts. It is reasonable to assume that each poster may have a few selected topics of interest, and there is a higher probability that they will post photos on such topics. However, Instagram does not provide a set of pre-defined topics or genres for photos. By viewing all tags added to photos by a poster, as a bag of words, therefore, I tried to identify latent topics of the poster. To do this, I first randomly selected 100K posters. I then applied a Latent Dirichlet Allocation (LDA) model (Blei et al., 2003), using Mallet (McCallum, 2002), an open-source machine learning toolkit, to identify a list of latent topics per poster.

Table 3-4: LDA-discovered topics in Instagram (N=100K).

ID	Topic	Tag examples
1	Nature	sky, nature, flowers, ocean, beach
2	Fashion/beauty	makeup, jewelry, model, fashion, beauty
3	Location/place/area	nyc, boston, spain, italy, brazil, home
4	Art/photos/design	photo, interior, architect, design, art
5	Holiday/vacation	party, holiday, vacation, friday, rest
6	Mood/emotion	love, cute, happy, smile, great, good
7	Social/people/family	family, girlfriend, boyfriend, gay, folks
8	Sports/activity	skateboarding, hiking, soccer, basketball
9	Entertainment	music, movie, pop, rock, song, play, star
10	Follow/shoutout/like	tagsforlike, followme, likes, shoutout
11	Food/drink	food, coffee, yummy, delicious, eat
12	Health/fitness	fitness, cleaneating, fit, yoga, workout
13	Animal	cat, kitty, instacat, pet, puppy, animal
14	Car/airplane	ford, Toyota, dodge, hotcars, bmw, truck
15	Travel	mytravelgram, trip, instatravel, traveling
16	Religion/belief	blessed, god, faith, truth, jesus, mind
17	Funny/quotes	lol, funny, jokes, quotes, saying, lmfao
18	Technology	samsung, galaxy, iphone, ipad, computer
19	Smoking	weedstagram, high, weed, dope, smoker
20	Apps/games/comics	instahub, webstagram, comics, gamer

As Mallet generated different topics for each execution, I ran Mallet 50 times to extract 100 well-presented topics. Mallet generates two types of outputs—a list of keywords for each topic and the ratio of each topic per poster. Because I found that there were some overlaps among the 100 topics, I categorized them by taking a bottom-up approach. First, to obtain ground-truth tag categories, I investigated a number of third-party websites that present a list of popular or hot tags in different time frames (i.e., daily, weekly, monthly) and finally chose two websites (i.e., tagsforlikes.com, tagstagram.com). Based on those categories, three human judges then inductively coded the types of topics and continued this process until all judges agreed. At the end, through this process, I was able to identify 20 mutual Instagram topics. Each poster had 20 topics with a different ratio depending on the tags added to the photos. Table 3-4 shows a list of final 20 topics with some tag examples. Lastly, to obtain the number of Likes per topic, I multiplied the ratio of each topic with the total number of Likes. This was a reasonable method to find the relationship between Likes and topics, because I found that the number of Likes tends to be evenly distributed over one's photos.

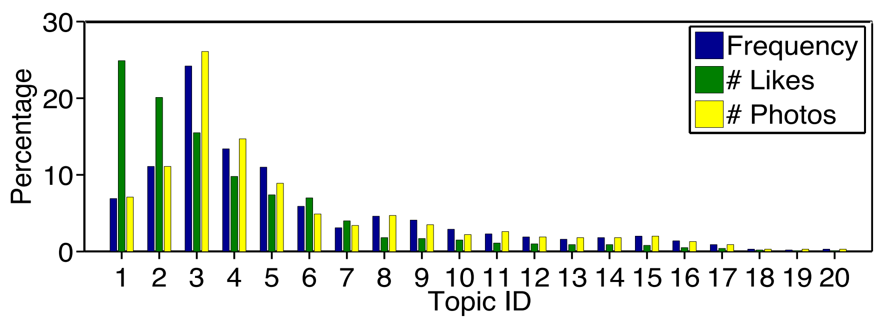


Figure 3-3: Ratio of frequency, # of Likes, and # of photos for the topics in Instagram (N=100K).

Figure 3-3 shows the ratio of frequency, the number of Likes, and the number of photos for each topic. First, regarding frequency, "Location/place/area" (3rd topic in Table 3-4) was the most frequent topic to be found in Instagram, followed by "Art/photos/design" (4th), “Fashion/beauty” (2nd), and “Holiday/vacation” (5th). I found that the tags in “Follow/shoutout/like” (10th) especially represented a unique and interesting culture in Instagram, as I can assume that posters with those tags in “Follow/shoutout/like” tend to desire to have their photos more widely searched and accessed by other users. Second, the most frequent topic is not necessarily the one receiving the most number of Likes. For example, the frequency in “Nature” (1st) was only 6%, but its number of Likes (24%) was the highest. “Location/place/area” (3rd) had the highest frequency (24%), but was not the highest regarding the number of Likes (15%). Lastly, unlike the Like results, the number of photos broadly showed a similar pattern to their frequency, which further implies that in general tags were quite well distributed across one’s photos.

As to the number of Likes and photos, most topics showed a small difference except for the first four topics. “Nature” (1st) received more Likes than photos, whereas “Location/place/area” (3rd) had more photos than Likes. Regarding the “Nature” (1st) topic, I speculate that there might be many high-quality photos showing the beauty of the nature that affect user behavior. That is, users who posted those photos might prioritize the quality of photos, but not necessarily their quantity, which might attract more users and make them to add Likes. “Location/place/area” (3rd) showed the highest number of photos, because the tags in this topic seem to describe a wide range of photos that are used together with many other topics. This perspective can be partly supported by a high frequency of their usage.

Overall, based on the dataset, it appears that the first five topics (IDs between 1 and 5) represent the main contents posted and shared by users in Instagram with respect to their frequency of usage, and the number of Likes and photos. This may not be generalized to the whole set of activities in Instagram. However, I believe that this result shows the connection between photo topics and Like activities in an online photo-sharing community.

Chapter 4

Studies on generation in social media

Perspectives on teens in social media

The notion of “digital natives” was first proposed by Prensky (2001), which describes a new generation who has spent their entire lives surrounded by technologies and tools of the digital age. This notion has sparked a wide range of debate. Much of the opposition argues that the so-called digital natives do not necessarily possess the natural fluency and technological skills that they are assumed to, nor are they necessarily more intensive users of digital media than many so-called “digital immigrants” — people of the older generation who transitions from traditional media to new (Thomas, 2011).

However, even though the term might exaggerate the inter-generational gap and overlook the intra-generational digital divide, research has shown that young people in general, especially teens, are highly tech-savvy (Kennedy et al., 2008). When it comes to social media contexts, studies show that teens tend to be early and fast adopters of newer or better online social space (Birnholtz, 2010). They also show that teens tend to use multiple social media sites and maintain different forms of communication (Quinn, Chen, & Mulvenna, 2011), and can quickly switch between different platforms to take advantage of their unique features (Quan-Haase & Young, 2010).

As a result of the digital proficiency and skillfulness, teens are likely to be active users of digital media. For instance, a national survey of teens in 2009 confirmed that age and Internet access are positively associated with digital literacy and Internet use (Livingstone & Helsper, 2009). Therefore, I hypothesized that teens would exhibit more behavioral activities in social media, and would be capable of utilizing more technological features afforded by the platform:

H1: Teens are more active users of social media than adults, because they will be engaged in more behavioral activities and utilize more technological affordances on the social media (in my case, Instagram).

Teens & other age groups on social media

Some of the differences in social media use between teens and adults can be explained by their different levels of digital literacy and perceived competence. However, as discussed above, this assumed generational digital divide is rather ambiguous in reality. As the design of social media interfaces has become increasingly intuitive and easy to use, adult users are quickly catching up in number when it comes to the use of some of the most popular social media websites, such as Facebook and Pinterest (Duggan & Brenner, 2013). If technological literacy is not the determining factor, what other factors might lead to the unique behavioral patterns of the different age groups?

Social media offers abundant opportunities for social connections and social interactions; therefore, it serves to provide a virtual "social context" (Adams & Marshall, 1996), an immediate social environment in which social and situational variables can greatly shape individual behavior. Therefore, individuals would behave in accordance with the social norms and in response to social influence they experience in social media. Previous developmental literature suggests that teens are particularly prone to such social influence. Consistent with such theoretical assertions, uses and gratification research (Papacharissi, Z. & Mendelson, 2011; Park, Kee, & Valenzuela, 2009) has also found that individuals use social media mainly for relationship maintenance, social surveillance, and social interaction, among other purposes (e.g., entertainment, self-status seeking, information seeking, etc.). However, especially for teens, communication with their peers emerges as the single most important motivation for SNS use (Barker, 2009). Ethnographic data have shown high teen engagement in the online socialization opportunities and social behaviors unique to the mediated environment (boyd, 2008). Teens also tend to maintain a social network with a large number of users (Pfeil, Arjan, & Zaphiris, 2009) and consider social media a place for self-representation (Jang et al., 2015; Livingstone, 2008), and for establishing their own identity (boyd, 2008).

In this regard, I hypothesized that teens would exhibit more social activities than adults in order to stay connected with their peer group members through various means that are unique to the social media site:

H2: Teens are more engaged with social interactions with other users than adults through communication features (e.g., Likes, comments, tags, etc.) offered by Instagram.

Overall, there have been a lot of research efforts on studying teens and generational perspectives in social media. However, relatively few studies have articulated differences from a large, data-driven longitudinal and comparative analysis. To fill this gap, this thesis introduces a new method and presents less explored aspects, and a comprehensive picture of teens' social media behaviors.

Chapter 5

Teens in Instagram

In this section, I report my analyses on the usage data from a total of 26,885 teens and adults. I first briefly summarize the two primary usage differences (i.e., teens tend to post less and be engaged more than adults) between two groups, and introduce and explain several factors that may influence those two findings. My overall analysis is guided by the following research questions.

RQ1: Do teens behave differently from adults in their Instagram activities? Are teens more active users than adults?

RQ2: If, indeed, an age difference is identified with respect to activities and engagement in social media, what are social, psychological or technological factors that may explain such behavioral differences?

Regarding RQ1, this study shows mixed findings. Comparatively, adult users are found to have posted more content with a higher level of topic diversity, whereas teens tend to be engaged in more interactions with their social networks through Likes, comments, and the expression of emotion and social interests.

Regarding RQ2, two hypotheses — *teens as digital natives* and *the need for social interactions* — based on previous literature and related work will be tested. I identify behavioral patterns which indicate that, while self-representation seems to be universally important for social media users irrespective of age, teens tend to show more behavioral activities in social media, manage their social media content to meet their social needs, and interact with more diverse users. Adults tend to focus on expressing their identity and engagement through content creation (i.e., adults tend to post more photos than teens), and interact with relatively smaller number and less diverse users.

User collection method

Classifying users to a specific age group is challenging. Most social media platforms, including Instagram, neither collect nor disclose users' age information. To address this challenge, I proposed a method that leverages two existing media contents (i.e., bio descriptions and profile images) with existing APIs (Jang et al., 2015). First, I applied textual pattern recognition algorithms to parse a list of patterns that specifically describe users' age in the bio (e.g., *"I am 17 years old," "I'm 23"*). Second, I used a facial recognition technique, Face++, to auto-detect the age information from people's profile images, which has been utilized and showed a high accuracy in another study (Bakhshi et al., 2014). Figure 5-1 illustrates the method and process of data collection.

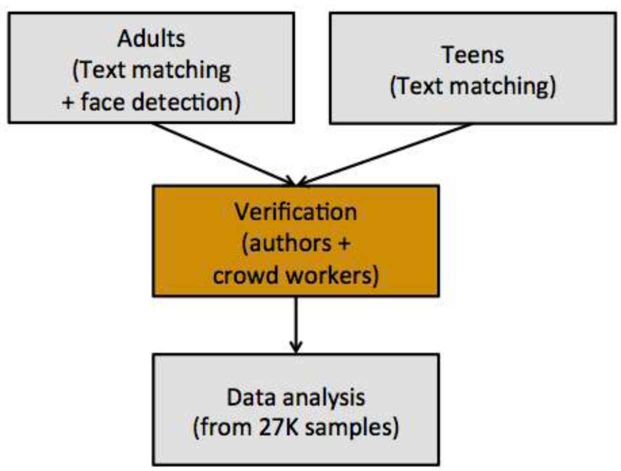


Figure 5-1. Data collection method and process (a mix of text matching, face detection, and crowd workers).

I defined my target user populations as follows (I intentionally had a five-year gap between two groups to minimize ambiguity in ages):

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

Based on the users' bio descriptions, I identified 13,533 teens and 8,596 adults. To fill the gap between two groups, then, I analyzed the rest with Face++ and added 4,756 more adults, resulting in a total of 26,885 teens and adults for the analysis. This number was after the manual verification of the age of all users to make sure the data accurately represented each group from a total of five human judges (i.e., two authors and three crowd workers in Amazon Mechanical Turk).

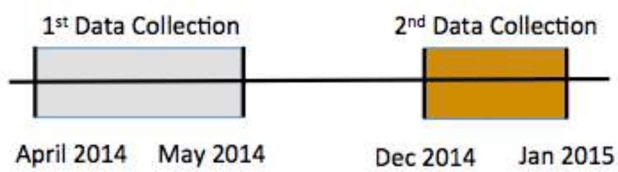


Figure 5-2. I found a trend from the 1st dataset and temporally extended it to create the 2nd dataset with the same number of users (13,533 teens and 13,352 adults).

To better see the patterns of usage and engagement over time in Instagram, I additionally collected the dataset from the same 26,885 users in 12-hour intervals over 12 days (from Dec. 26, 2014 to Jan. 6, 2015). Figure 5-2 illustrates the construction of two datasets. For the analysis, I calculated the delta of photo counts in every two consecutive time slots and checked the total number of photos that users have posted, the number of users who added photos, and the number of users who removed their photos in 12 hours.

In order to protect the privacy and confidentiality of the Instagram users in the sample, privacy-preserving measures were taken throughout different stages of this study. Specifically, during the process of manual age verification, I removed identifiable and sensitive information (i.e. name, ID, and email address) from the profiles and photos before they were presented to crowd workers. Moreover, during data analysis, I removed all personally identifiable information (except for age), and aggregated and analyzed the data at a group level.

Results

In this section, I report the analyses on the usage data from a total of 26,885 teens and adults. I first briefly summarize the two primary usage differences (i.e., teens tend to post less and be engaged more than adults) between two groups, and introduce and explain several factors that may influence those two findings.

Teens' behavioral differences

My analysis on the usage of Instagram from all users showed that *teens tend to post fewer Photos but show more activities in Liking, Tagging, and Commenting* (Table 5-1). As all variables show a long-tailed distribution, I used the median value for the analysis.

Table 5-1: Summary of activities by two groups. Teens tend to post less but be engaged more in other activities than adults.

	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

I also calculated the ratio of Likes, Tags, and Comments to Photos and found that teens are likely to receive more Likes (teens: 56.10; adults: 40.03; I used eta-square (η^2) for the effect size: 0.09), add more tags (teens: 6.34; adults: 4.70; η^2: 0.01), and have more comments (teens: 2.52; adults: 1.06; η^2: 0.07) per photo than adults (statistically, all showed significant differences; $p < 0.001$). This result indicates that the way of using and engaging in Instagram between teens and adults is quite different. In the following sections, I investigated several factors that might lead to these results.

Factors of behavioral differences (1): teens have fewer photos in Instagram

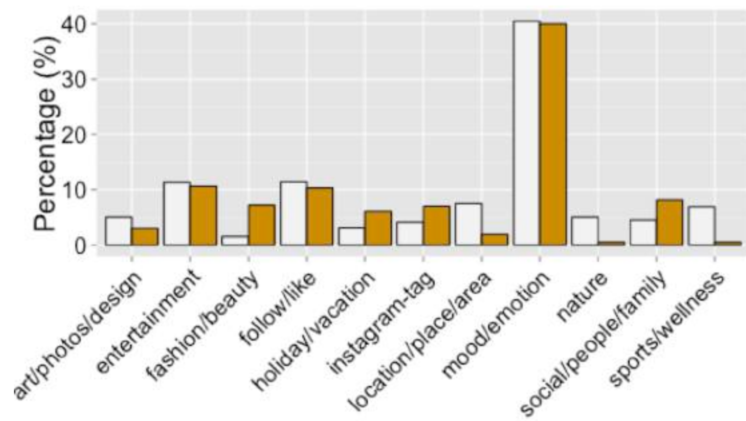
Lack of topic diversity of photos

I first examined the list of topics that are presented in teens' and adults' photos. I assumed that, for teens, activities and topics of photos might be limited, because they are financially or culturally dependent on their parents to venture outside of their daily activities compared to adults. To test the assumption, I used tags in both posted and removed photos in order to find the topics of photos that users have or used to have (but removed).

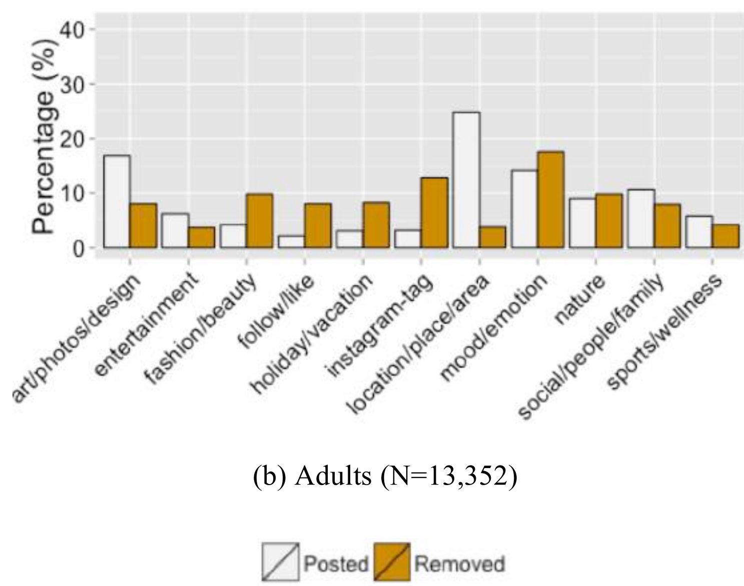
Table 5-2: LDA-discovered topics from all users (N=26,885). Tags were used for topic discovery.

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

I identified latent topics from the tags of users' photos through the LDA analysis (Blei, Ng., & Jordan, 2003) using Mallet (McCallum, 2002). I used tags same as the last section. I also obtained a list of ground-truth tag topics from two popular websites (i.e., tagsforlikes.com and tagstragram.com). The types of photo topics were manually coded from Mallet's output into those topics. Table 5-2 summarizes the 11 topics extracted from the dataset. I then calculated the percentage of topics from posted and removed photos for each group, as presented in Figure 5-3.



(a) Teens (N=13,533)



(b) Adults (N=13,352)

Figure 5-3: Percentage of the posted and removed photos based on LDA-discovered topics (x-axis), where N=26,885. Teens show a very high result in the mood/emotion topic both for posted and removed photos whereas adults show more diverse topics.

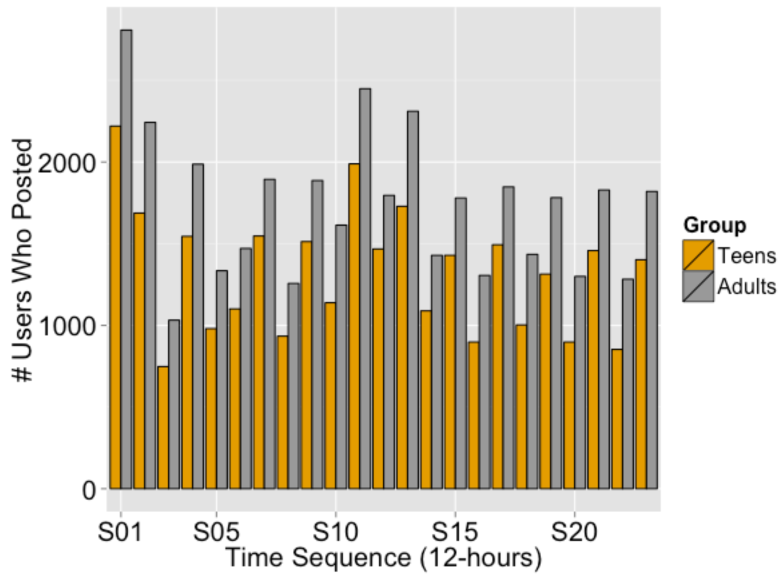
Figure 5-3 shows a clear difference between two groups in terms of topic types. On the one hand, for teens, more than half of posted and removed photos were in “Mood/emotion” and “Follow/Like.” These topics are not necessarily tied to the content of photos but rather describe one's emotional status or intention to have more followers. In addition, topics of posted and removed photos for teens are highly correlated ($r = 0.92$, $p < 0.001$), indicating teens show quite similar patterns when managing their photos. On the other hand, adults showed a high ratio in more diverse topics from their posted photos, including "Arts/photos/design," "Locations," "Mood/emotion," "Nature," and "Social/people."

Unlike popular topics presented in teens' photos, these topics imply more diverse content in the photos, such as photos that depict different facets of cities and countries around the world, photos of arts and design (some of them were taken professionally), photos of a variety of people, and so on. Similarly, adults present quite diverse topics from removed photos, and, unlike teens, topics from posted and removed photos do not correlate with each other ($r = 0.09$, $p = 0.77$). In summary, it appears that teens' posted and removed photos presented a lack of diversity compared to adults' ones.

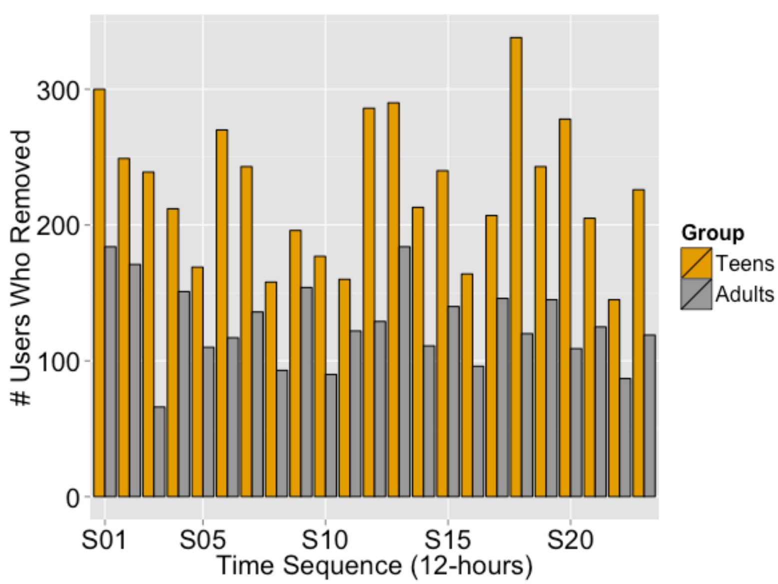
Post fewer and remove more photos

In addition to the topics of photos, I used the data showing the temporal usage reports (collected in the second phase) to measure the number of users who posted or removed photos, and that of posted or removed photos. For removed photos, I checked if each individual photo still existed by comparing a list of photos every 12-hours.

Figures 5-4 and 5-5 show the number of photos posted or removed and the number of users who posted or removed photos over time, respectively. From almost the same number of users in each group, the results show that fewer teens tend to post photos ($t(42) = -3.89$, $p < 0.001$), but more teens tend to remove photos than adults ($t(42) = 8.01$, $p < 0.001$). Regarding the total number of photos posted and added, teens posted fewer photos ($t(42) = -3.76$, $p < 0.001$) and removed more photos than adults ($t(42) = 6.14$, $p < 0.001$).

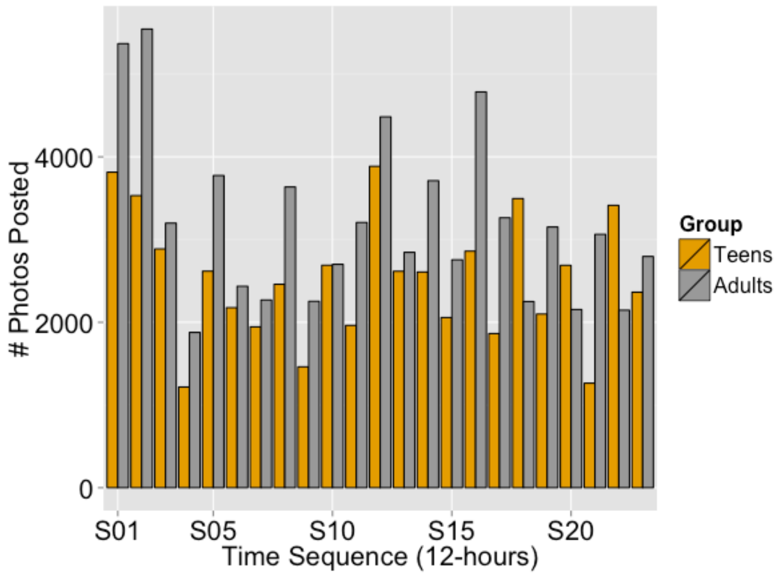


(a) # Users who posted photos

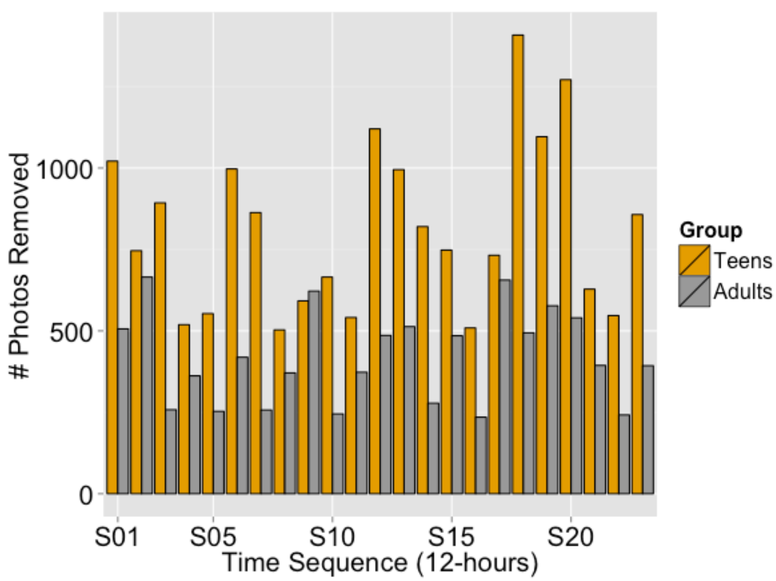


(b) # Users who removed photos

Figure 5-4: Number of users who (a) posted or (b) removed photos over 12 days (N=26,885). More adults post and more teens remove photos.



(a) # Photos posted



(b) # Photos removed

Figure 5-5: Number of photos (a) posted and (b) removed over 12 days (N=26,885). More photos posted by adults and more photos removed by teens.

Remove photos with few Likes

Prior research has found that many teens tend to manipulate their photo content to receive as many Likes as possible or sometimes remove some photos that have received too few Likes (boyd, 2008; Madden et al., 2013). Because there has not been any attempt to examine this phenomenon through a large-dataset analysis, I measured whether the usage data from the sample showed a similar perspective, which also supports the idea of teens having fewer photos.

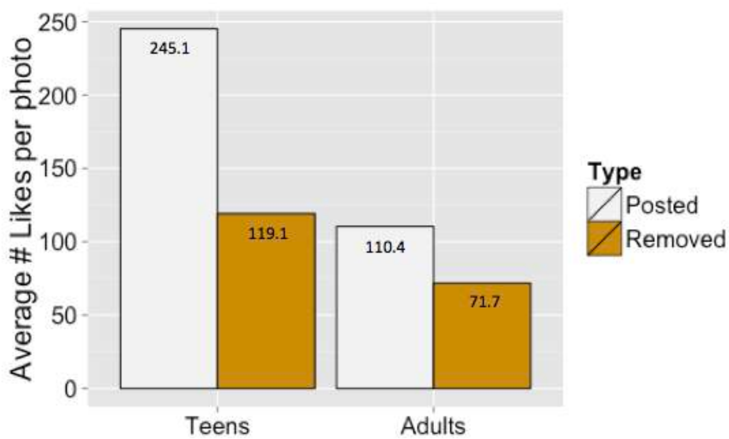


Figure 5-6: Average number of Likes per photo for posted and removed photos (N=26,885). Removed photos from adults have 35.0% fewer Likes than posted photos, whereas teens show 51.4% fewer Likes. Teens remove more photos that have fewer Likes than adults.

I first measured the number of Likes that the removed photos received and compared it to the number of Likes that all posted photos garner. As I had the usage data 12-days in a row, I could calculate how many Likes removed photos received. I then checked if teens removed photos that had relatively fewer Likes. Figure 5-6 shows the result of the differences. The average number of Likes per posted photo is 245.1 for teens and 110.4 for adults. The average of Likes per removed photo is 119.1 for teens and 71.7 for adults. Then adults' removed photos have 35.0

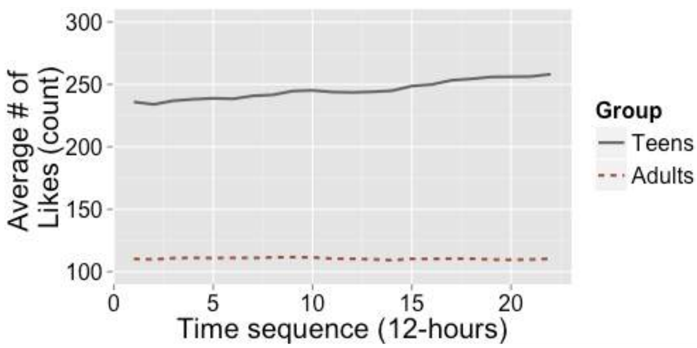
percent fewer Likes than the posted (and kept) photos, whereas teens' removed photos have 51.4 percent fewer Likes than their posted (and kept) photos ($t(44) = 7.08$, $p < 0.01$). This result indicates that both user groups tend to remove photos that have fewer Likes than their overall photos, but teens show a larger difference compared to their posted photos.

Factors of behavioral differences (2): teens engage more in Instagram

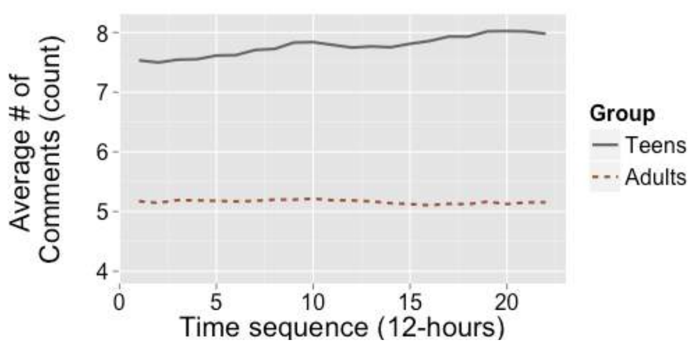
To understand the level of engagement, I examined four aspects of teens and adults including: (1) how many Likes and comments they have had over time, which implies a level of activities in Liking and commenting; (2) how fast they replied to other users' comments added to their photos, which implies a level of one's interest in interacting with other users; (3) how they engaged with other users through comments, which implies a level of their engagement; and (4) what they wrote in their comments, which implies their intention through commenting.

Teens have more Likes and comments over time

I measured the average number of Likes and comments teens and adults have had over time. Figure 5-7 illustrates the average number of Likes and photos over 12 days. Not surprisingly, teens showed the higher number of Likes and comments every day than adults ($p < 0.0001$). However, a more interesting aspect is that teens presented an overall increase, especially in Likes. In addition, when I consider this as a cumulative result, a total number of Likes for teens will be much higher than adults, which also implies high engagement in Liking. On the other hand, for adults, the average number of Likes and that of comments do not seem to be changed a lot. This implies that adults are likely to receive the similar number of Likes, even after they add new photos (i.e., in the previous section, I found that more adults post more photos than teens as shown in Figures 5-4 and 5-5). This further means that adding more photos does not necessarily lead to having more Likes or comments in adults' case. Figure 5-7 does not show a saturation point (i.e., no increase after reaching the certain number of Likes) for teens. However, for adults, the number seems to reach the threshold of having around 110 Likes and 5 comments. Overall, this finding supports well the idea of teens engaging more in Instagram activities than adults.



(a) Likes



(b) Comments

Figure 5-7: The average number of Likes and comments in every 12 hours (N=26,885). Teens show the steady increase over time for both Likes and comments while adults remain flat.

Teens reply to others' comments more quickly

Adding user's name right after the "@" symbol has been widely used in social media for replying to another user and helping establish a language for communicating. I can think about a scenario where an original photo poster, *@robinson*, checked one comment (e.g., *"Nice pic, where did you take it?"*) added to his photo by another user, *@johndoe*. Then, *@robinson* added a new comment (e.g., *"@johndoe, I took this photo when I visited New York"*) and mentioned *@johndoe* in his comment.

Based on the scenario above, I measured how quickly the original photo posters replied to other users' comments that were added to their photos. As shown in Figure 5-8, I found that teens replied to the previous comments from other users in around 7.2 minutes, which is significantly shorter than adults who replied in around 30.0 minutes.

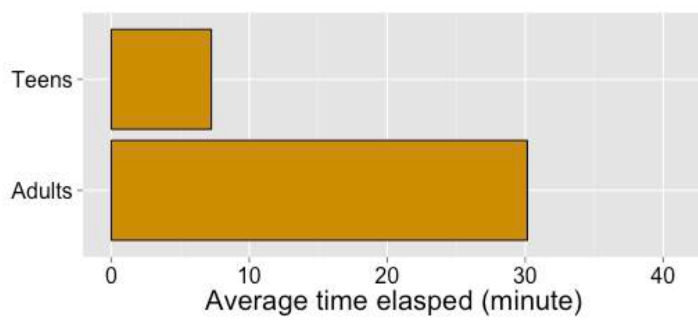


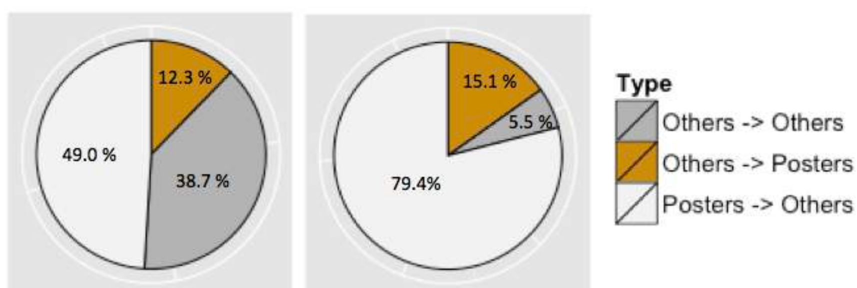
Figure 5-8: Avg. time elapsed when the original photo posters (N=26,885) commented and mentioned *@name_of_previous_commenter* right after *@name_of_previous_commenter*'s comment. Teens tend to reply more quickly than adults.

Teens engage with more diverse users through commenting activities

To measure a number of other users that the sample users are engaging with through their comments, I first calculated the ratio of the number of comments with *@others_username* to that of all comments for each group. I found that teens showed a higher result (45.2%) than adults (34.1%). This indicates that more teens are adding *@username* in their comments in order to, for example, call other users, or start and maintain conversations.

I further examined additional types of commenting with respect to two dimensions: (1) "users in the comments" — those who added comments and those who were mentioned in the comments — and (2) "user types" — original photo posters and other users. By combining these two dimensions, then, I derived three directional aspects of commenting with @: (1) original photo posters mentioned other users (posters → others); (2) other users mentioned the original photo posters (others → posters); and (3) other users mentioned other users (others → others), excluding the original photo posters, in their comments.

Figure 5-9 (below) shows the break-downs of three types. First, teens have fewer cases (49.0%), where the original photo posters commented *@others* (posters → others) in their comments, than adults (79.4%). Second, teens have fewer cases (12.3%), where other users mentioned *@photo_posters* in their comments (others → posters), than adults (15.1%). Third, teens show more cases (38.7%) where other users mentioned *@others* in their comments (others → others) than adults (5.5%).



(a) Teens (N=13,533)　　(b) Adults (N=13,352)

Figure 5-9: Results of three types of adding comments with @. Teens show a higher result in "others→others" than adults, and adults show a higher result in "posters→others" than teens.

When I consider these results together, it was found that most comments (79.4% + 15.1% = 94.5%) in adults' photos are associated with the adult photo posters, whereas more than half of the comments (49.0% + 12.3% = 61.3%) in teens' photos are associated with teen photo posters.

It is interesting to see that many other users who commented on teens' photos mentioned other "third users" in their comments (38.7%), whereas only very few cases of others → others are observed (5.5%) in adults' photos.

To gain a more concrete idea of patterns of commenting, I further measured the percentage of (non-overlapping) unique users who were mentioned in photo poster's comments, and the percentage of unique users who mentioned the photo poster in their comments.

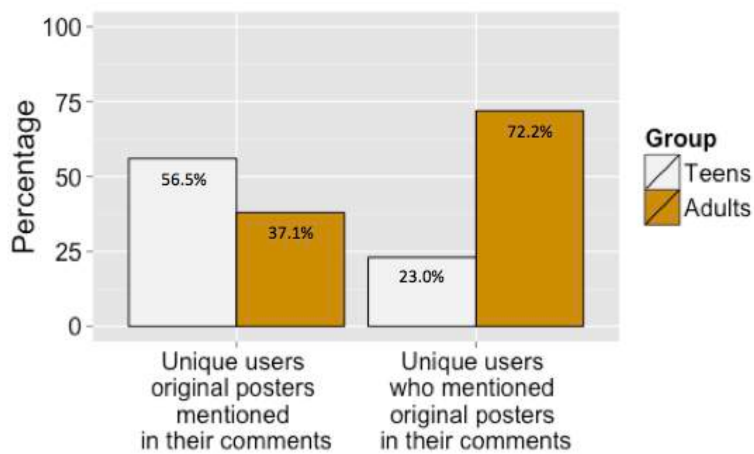


Figure 5-10: Percentage of (non-overlapping) unique users whom photo posters mentioned in their comments and unique users who mentioned the photo posters in their comments with @. For teens, (1) the original photo posters mentioned more unique users in their comments and (2) other users who added comments to teens' photos mentioned more third users. However, these were opposite in adults' comments.

Figure 5-10 shows the results with two interesting insights. First, teens mentioned other users in their comments (56.5%) more than adults (37.1%). Second, a majority (72.2%) of the comments on adults' photos were directed toward the original photo posters by mentioning their usernames, whereas less than a quarter (23.0%) of the comments on teens' photos were directed toward the original photo posters. Instead, 77.0% of the comments on teens' posts mentioned other people's usernames.

Given that comments and mentions can reach users instantly through push notifications (see Figure 5-11) and are therefore effective communication tools on Instagram, this result shows the *different communication patterns* between teens and adults. Teens are known to be highly active in being connected with others (Grinter, Palen, & Eldridge, 2006; boyd, 2008) through texting and social media (Pater, Miller, & Mynatt, 2015) and responding quickly (Figure 5-8). Thus, the comparison im Figure 5-10 reveals that teens tend to use the comment section to reach out to and interact with a more diverse and bigger network of users, while adults tend to use comments and mentions to have a more direct and interpersonal interaction (with fewer other users), showing more person-to-person or individual-oriented interactions.

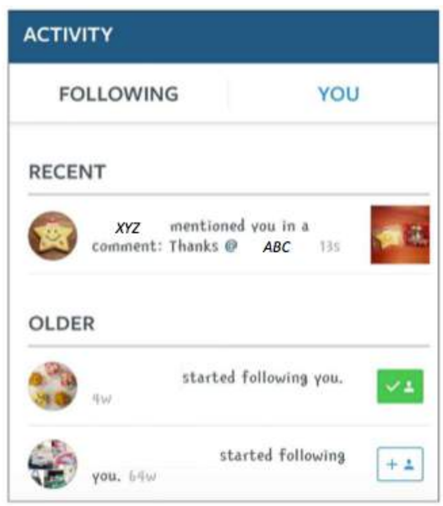


Figure 5-11. Instagram interface that shows one's activity. When a user (named *XYZ; anonymized*) adds @*ABC (anonymized*) in his comment, a notification will be sent to *ABC* as well as added to her activity page, indicating her name was mentioned in @*XYZ's* comment.

Teens post emotional social interests

Lastly, I measured the words in the comments with @ by teens and adults as the response to others' comments. I used Linguistic Inquiry and Word Count (LIWC) (Pennebaker & Francis,

1999) in order to parse words representing tones and psychological components in the comments. I first randomly chose a total of around 6,000 teens and adults and compared word count (i.e., related to talkativeness and verbal fluency), the number of words longer than six letters (i.e., the higher it is, the less emotional and connected), and a level of social interests and that of emotionality.

Table 5-3: Analysis on comments from randomly selected teens (N=2,927) and adults (N=2,928) using LIWC. Teens show a higher result in social and emotional connection in their comments. Eta-square (η^2) was used for the effect size.

	Teens	**Adults**	**Effect size (η^2)**
Word count*	86.4	191.9	0.04
Words > 6 letters*	9.1	10.4	0.02
Social interests*	13.7	11.2	0.01
Emotionality*	38.0	31.0	0.07

*$p < 0.001$.

Table 5-3 summarizes the result. The word count and words longer than 6 letters for teens were lower than for adults. However, social interests and emotionality were higher for teens than for adults. It shows that although adults add longer comments, their comments are less emotional and oftentimes psychologically distant. Teens' comments were shorter but more emotional and embodied social aspects. Teens might spend less time on commenting because they have shorter texts. This perhaps implies that promptness is an important factor for teens when interacting with others as shown in Figure 5-8.

Chapter 6

Discussion

I have presented the in-depth analysis on the social media usage of two different age groups – teens and adults. I primarily focused on investigating the factors that may affect the earlier findings where teens tend to have fewer photos but be more engaged in Instagram. I attempted to detail the reasons behind these trends based on the theoretical foundations and the analysis on the large datasets collected from a total of 26,885 teens and adults through mixed methods (i.e., text matching, face detection, and crowd workers). In this section, I first summarize two trends and the underlying factors, which I assumed and found through the analysis, and then discuss those factors within a lens of my hypotheses. I finally discuss study limitations and future work.

Two trends and the corresponding factors identified

For the first trend, where teens have fewer photos than adults, I have identified the following insights that explain it:

- Having fewer topics to post seems to lead teens to post fewer photos over time. Teens are likely to add/remove similar or quite limited topics that mostly describe their emotional status or intention to have more followers, while adults showed more diversity.
- Teens tend to add fewer photos and remove more photos than adults based on the analysis on the temporal usage dataset.
- Teens tend to remove photos that have relatively fewer Likes compared to adults.

For the second trend, where teens engage more than adults, I have discovered the following insights:

- Teens have more Likes and comments per photo than adults, and my analysis from the temporal usage dataset shows that teens are likely to receive more of them.
- Teens tend to respond more quickly to others' comments added to their photos than adults.
- Teens tend to add more comments with @username and have more non-overlapping (unique) users mentioned in their comments. Users who added comments to teens' photos also mentioned many diverse third-users in their comments. Thus, given the fact that users will receive a push notification whenever there is a new message added to their photos or comments, it is likely that the original photo posters (teens) will see and be aware of many usernames mentioned in the comments. Teens also add more comments that show social interests and emotionality. On the other hand, adults tend to post fewer comments with @ and mention fewer unique users in their comments. Users who commented on adults' photos mentioned the original photo posters more than teens.

Theoretical and practical interpretation and implication

Hypothesis on digital natives

First of all, H1 (digital native hypothesis) was partially supported. Teens in Instagram were fairly active in all activity categories but were not necessarily more active than adult users in all aspects. In particular, they were found to create less content than adult users, despite their high engagement in commenting, Liking and tagging activities.

This finding has several implications. First, it shows that social media engagement is a multi-facet concept that encompasses not only content creation but also social interactions in various means. Second, it indirectly supports the assumption of tech-savvy teens (Birnholtz, 2010; Kennedy et al., 2008), in the sense that teen users were more likely to utilize the diverse features afforded by the interface for social networking purposes. Both hashtags and Likes are unique features of the new social media, which may be unfamiliar to some of the adult users. However, teens in my sample effectively utilized such features for proactive socialization.

Third, while examining teens' activities in Instagram, I identified an interesting pattern that they tend to manage their personal profile through content removal. This could also be an indicator of their skilled use of social media, where I found a possible link to a privacy aspect. For instance, teen privacy research has suggested that technology-savvy and -native teenagers would limit or remove their online postings (often after the fact) as a privacy protection mechanism rather than limiting their overall online activities and information revelation (Jia et al., 2015). Similarly, based on the survey result out of 622 teens, research has found that 62 percent of teens (382) deleted or edited their content posted in the past as a way of their privacy strategy (Feng & Xie, 2014). This unique strategy shows how today's teens manage their online content in different ways than older generations. In addition, content deletion could be a novel way to manage teens' online self-representation, and this is also related to my second hypothesis.

Hypothesis on social interactions

Secondly, H2 (social interaction hypothesis) was supported. My findings showed that social interaction was the primary motive for teens and had significantly shaped their behaviors in Instagram. Not only did teens receive more Likes and comments than adults — and following the social rule of reciprocity (Falk & Fischbacher, 2006), I could assume that they left more Likes and comments on other users' profiles prior and/or in return — but their content deletion appeared to be associated with a lack of Likes. Teens' content management strategy appears to be for the purpose of self-presentation: compared to adults, teenagers especially may want to display the “popular self” (Ito et al., 2008). For them, they would think only keeping the most Liked posts can help create a perception that the profile is popular. This result shows a strong empirical and longitudinal field evidence for the previously established relationship between online self-presentation and psychological factors such as self-esteem and narcissism (Mehdizadeh, 2010).

My analyses showed that adults appeared to create more original content and to have kept more user-generated content (UGC) than teens. This finding shows some interesting perspectives. On the one hand, adults may have access to more resources and life experiences, which serve well as their source of content creation, while such resources and experiences are lacking for teens. On the other hand, existing research has suggested that, different from consumption and participation in UGC, the production of UGC is primarily driven by the needs of self-actualization (Shao, 2009).

Such needs can be more salient for adults as they have well-established identities and confidence in voicing their identity. Shao pointed out that participation in UGC, in forms such as commenting and Liking, is associated with social needs. Given that social needs are more salient for teen social media users, it would be reasonable to see the result where teens were more engaged in these social interactions than content creation. From the social capital perspective, despite the fact many scholars believe the active use of SNS is more effective in achieving social capital, some research indicates that passive use (e.g., Liking, commenting, or just lurking) of SNS can also function as a form of social investment and therefore contributes to social capital (Burke, Kraut, & Marlow, 2011).

Another insight can be revealed from the finding in which teens were engaged in more limited topics when they did create content in Instagram. The limited topic diversity might be explained by the hypothesis of online environment as an “echo chamber,” referring to a situation in which information, ideas, or beliefs are amplified or reinforced by transmission and repetition inside an enclosed system (Sunstein, 2002). In the context of this study, the highly personalized content consumption enabled by online services allow users, young adults and adolescents in particular, to select only the content that they are interested in and the opinions that they agree with. Such selective exposure may lead to their limited scope of interest and topic diversity.

Lastly, my findings about comments and mentions in comments revealed the different strategies that teens and adults utilize to interact with their social networks. The fact that the mentions in comments on adults' profiles were used more frequently for direct communication with the original photo posters shows the adult users' preference for having or maintaining close, interpersonal interactions. Teens, however, tend to use the comment space to reach more and other users. They also responded more quickly to these comments through the Instagram notifications. This shows that teens maintained wider and more timely interactions with a large network of people. The social interaction that takes place in the comment space may have compensated for the limited content posted onto Instagram as well. This also in part affects the increase in Liking that teens showed compared to adults as teens make more new friends while adults seem to be more interested in interacting with established friend groups.

In summary, my data-driven and comparative analysis unearths several new and unique insights on teens and their behaviors in social media. At the same time, the analysis substantiates the idea that teens leverage social media primarily as a "conversation space" (boyd, 2008) and use many features the platform provides in order to create connections and facilitate conversations and interactions (Ito et al., 2008). Teens engage in social media not only because they are well aware of the intention of those activities, but also because they are familiar with technology use and the "tagging culture" in online space, which reinforces their social practice (Adams & Marshall, 1996).

Practical interpretations and implications

Along with many theoretical insights I had from my results, there are some practical interpretations and implications especially about the design of social media sites (in my case, Instagram).

There is a design opportunity where social media sites can provide users with a summary of their usage reports. Many activity variables that are used for the analysis in this paper can be considered including the number of photos, Likes, comments, tags that users added, photo topics identified from the system, most popular photos based on the total number of Likes and comments, and so on.

Then the social media sites can leverage this design idea to provide one with a recommendation of other users who have shown similar activities or photos. The current design of Instagram provides some recommendation features (e.g., discover new people, get trending topics, etc.), which is somewhat limited to one's followship or location information, or a search feature, where users have to explicitly search for other user's names or specific tags. Having an activity-, interest-, and content-based recommendation feature will create more interactive social space, which is beyond the one with simply one's followers/followings or location. As teens and adults show distinctive usage differences in Instagarm, this feature will give them chances to discover, meet and interact with new people who show similar interests and activities and/or are in the similar age. For example, teens and adults may find a list of recommended users in their age more interesting and meaningful and want to check their photos, follow them, and so on, because they may have more personal connections to peers.

However, there also exists some privacy concern where users can be identified by their usage activities and recommended to others. Even if the data being gathered and analyzed are publicly available and accessible, users usually have no way to know about whether their data are used in research or about how to opt out. As boyd and Crawford suggest, social media scholars need to be aware of the "considerable difference between being in public and being public" (boyd & Crawford, 2012), and should therefore carefully consider privacy and ethical implications when collecting and analyzing publicly available data. Especially if the data concern teenage users, as in this study, it would be important to take the best measures to minimize potential risk and harm, remove personally identifiable information, aggregate and analyze data at a group level, and adopt systematic data management strategies to ensure data security. Thus, there should be a careful design of supporting this new feature; for example, allowing users to control the visibility of some of their usage reports.

Limitations and future work

Although I presented a number of insights, I acknowledge some limitations that can be handled in future studies.

First, errors may exist in the detection of age information even if I manually verified them. Many users provide additional social media links (e.g., Facebook, Twitter, etc.) in their profiles that I can leverage. Future studies that apply my method should obtain and corroborate additional age information from those sites.

Second, the age information auto-detected from users' bios or profile images could be incorrect when users have not updated them for a long period. This could affect the analysis of behavior differences by age. A possible remedy is, for instance, to double-check users' age information by comparing a user's "selfie" photos with the user's profile photos. However, a further study to validate its accuracy will be necessary.

Third, the results from my dataset may not represent the whole social media platforms and may only be limited to teens and adults in Instagram (Ruths & Pfeffer, 2014). I plan to extend this study to other social media sites (e.g., Facebook, Flickr, etc.) to validate my method and compare results.

Fourth, my quantitative analysis opens many opportunities for follow-up qualitative studies to derive additional insights on teens and their behaviors in social media; for example, I are currently investigating the following RQs: *"What are the similar and different motivational aspects in Liking, tagging, and commenting between teens and adults?"* and *"To what extent do teens and adults consider social media as a conversation space and do they take any strategic approaches?"*

Chapter 7

Conclusion

This paper contributes to deeper analyses on age differences in Instagram, more broadly in social media. Based on diverse and data-driven comparative analysis techniques, I first explored and articulated the activities in Instagram according to three research perspectives; structure, influence, and context. I found that the activities in Instagram are not limited to the follow-relationship. I also found that five other Instagram elements influence the number of Likes received to different extents. In addition, using an LDA-based tag analysis, I identified 20 latent topics, prevalent among tags added to photos, and presented top 5 topics in Instagram.

In addition, I tested my hypotheses developed through the lenses of social cognition, developmental strategy, and human-computer interaction in order to explain how age factors in social media behaviors. This study results highlight that: (1) teens post fewer photos but remove more photos; (2) teens' removed photos tend to have relatively fewer Likes; (3) diversity of topics in teens' photos is more limited; (4) teens engage more in Liking and commenting, by presenting higher engagement with more diverse users, quicker responses to others' comments, and more comments that have emotional words and social interests, than adults. The behavioral patterns identified in the data analysis show the overall age difference in online communication strategies such that teens and adults adopt to meet their social and self-expression needs and to accommodate with their technological skills and preferences. Overall, this study results provide a number of new and interesting insights as well as some guidelines for ongoing research studies in social media.

References

Adams, G. R. & Marshall, S. K. (1996). A developmental social psychology of identity: Understanding the person-in-context. *Journal of adolescence*, 19(5), 429-442.

Agosto, D. E., Abbas, J., & Naughton, R. (2012). Relationships and social rules: Teens' social network and other ICT selection practices. *Journal of the Association for Information Science and Technology*, 63(6), 1108-1124.

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. *Proceedings of the International Conference on Human Factors in Computing Systems*, ACM, 965-974.

Barker, V. (2009). Older adolescents' motivations for social network site use: The influence of gender, group identity, and collective self-esteem. *CyberPsychology & Behavior*, 12(2), 209-213.

Birnholtz, J. (2010). Adopt, adapt, abandon: Understanding why some young adults start, and then stop, using instant messaging. *Computers in Human Behavior*, 26(6), 1427-1433

Blei, D. M., Ng., A. Y., & Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.

boyd, d. 2008. 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*, 119-142.

boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.

Burke, M., Kraut, R., & Marlow, C. (2011). Social capital on Facebook: Differentiating uses and users. *Proceedings of the International Conference on Human Factors in Computing Systems*, ACM, 571-580.

Chen, Y., Chuang, C., & Chiu, Y. (2014). Community detection based on social interactions in a social network. *Journal of the Association for Information Science and Technology*, 65(3), 539-550.

Davison, R.M., Ou, C., Martinsons, M. G., Zhao, A. Y., & Du, R. (2014). The Communicative Ecology of Web 2.0 at Work: Social Networking in the Workspace. *Journal of the Association for Information Science and Technology*, 65(10), 2035-2047.

Duggan, M. & Brenner, J. (2013). The Demographics of Social Media Users. *Pew Research Center's Internet & American Life Project*.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication*, 12, 1143–1168.

Falk, A. & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293-315.

Feng, Y. & Xie, W. (2014). Teens' Concern for Privacy When Using Social Networking Sites: An Analysis of Socialization Agents and Relationships with Privacy-Protecting Behaviors. *Computers in Human Behavior*, 33, 153-162.

Gilbert, E., Bakhshi, S., Chang, S., & Terveen, L. (2013). "I Need to Try This!": A Statistical Overview of Pinterest. *Proceedings of the International Conference on Human Factors in Computing Systems*, ACM, 2427-2436.

Grinter, R.E., Palen, L., & Eldridge, M. (2006). Chatting with teenagers: Considering the place of chat technologies in teen life. *ACM Transactions on Computer-Human Interactions*, 13(4), 423-447.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an Imagined Community. *Journal of American Behavioral Scientist*, 55(10), 1294-1318.

Haferkamp, N., Eimler S.C., Papadakis, A.M., & Kruck, J.V. (2012). Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites. *Cyberpsychology behavior and social networking*, 15(2), 91-98.

Han, K., Jang, J., & Lee, D. (2015). Exploring Tag-based Like Networks. *Proceedings of the International Conference on Human Factors in Computing Systems*, ACM, 1941-1946.

Hawn, C. (2009). Take Two Aspirin And Tweet Me In the Morning: How Twitter, Facebook, And Other Social Media Are Reshaping Health Care. *Health Affairs*, 28(2), 361-368.

Hollenstein, L. & Purves, R.S. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, 1(1), 21-48.

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What I Instagram: A first analysis of Instagram photo content and user types. *Proceedings of the International Conference on Web and Social Media*, AAAI.

Huberman, B.A., Romero, D.M., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday*, *Peer-Reviewed J. on the Internet*, 14(5).

Ito, M., Horst, H., Bittanti, M., boyd, d., Herr-Stephenson, B., & Lange, P., et al. (2008). Living and learning with new media: Summary of findings from the Digital Youth Project. *MacArthur Foundation Reports on Digital Media and Learning*.

Jang, J., Han, K., Shih, P. C., & Lee, D. (2015). Generation Like: Comparative Characteristics in Instagram. *Proceedings of the International Conference on Human Factors in Computing Systems*, ACM, 4039-4042.

Jang, J., Han, K., & Lee. D. (2015). No Reciprocity in "Liking" Photos: Analyzing Like Activities in Instagram. *Proceedings of the International Conference on Hypertext and Social Media*, ACM, Cyprus, September 2015.

Jia, H., Wisniewski, P., Xu, H., Rosson, M.B., & Carroll, J.M. (2015). Risk-taking as a learning process for shaping teen's online information privacy behaviors. *Proceedings of the International Conference on Computer-Supported Cooperative Work and Social Computing*, ACM, 583-599.

Kennedy, G. E., Judd, T. S., Churchward, A., Gray, K., & Krause, K. L. (2008). First year students' experiences with technology: Are they really digital natives?. *Australasian journal of educational technology*, 24 (1), 108-122.

Lampe, C., Ellison, N., & Steinfield, C. (2007). A Familiar Face(book): Profile Elements as Signals in an Online Social Network. *Proceedings of the International Conference on Human Factors in Computing Systems*, 435-444.

Livingstone, S. (2008). Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. *New media & society*, 10(3), 393-411.

Livingstone, S., & Helsper, E. (2009). Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy. *New Media & Society*, 12 (2), 309-329.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). Teens, Social Media, and Privacy. *Pew Research Center's Internet & American Life Project*.

McCallum, A. K. (2002). *"Mallet: A Machine Learning for Language Toolkit"*.

Mehdizadeh, S. (2010). Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking*, 13(4), 357-364.

Moran, M., Seaman, J., & Kane, H.T. (2011). Teaching, Learning, and Sharing: How Today's Higher Education Faculty Use Social Media for Work and for Play. *Pearson Learning Solutions*.

Muscanell, N. L. & Guadagno, R. E. (2011). Make new friends or keep the old: Gender and personality differences in social networking use. *Journal of Computers in Human Behavior*, 28(1), 107-112.

Ong, E., Ang, R., Ho, J., et al. (2011). Narcissism, extraversion and adolescents' self-presentation on Facebook. *Journal of Personality and Individual Differences*, 50 (2), 180-185.

Ottoni, R., Pesce, J., Casas, D., Franciscani, G., Meira, W., Kumaraguru, P., & Almeida, V. (2013). Ladies First: Analyzing Gender Roles and Behaviors in Pinterst. *Proceedings of the International Conference on Web and Social Media*, AAAI.

Papacharissi, Z. & Mendelson, A. (2011). Toward a new(er) sociability: Uses, gratifications and social capital on Facebook. in Papathanassopoulos, S. ed. *Media perspectives for the 21st century*, Routledge, New York, 212-230

Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *CyberPsychology & Behavior*, 12(6), 729-733.

Pater, J. A., Miller, A. D., & Mynatt, E. D. (2015, April). This Digital Life: A Neighborhood-Based Study of Adolescents' Lives Online. *Proceedings of the International Conference on Human Factors in Computing Systems*, ACM, 2305-2314.

Pennebaker, J. W. & Francis, M. E. (1999). *Linguistic inquiry and word count (LIWC)*. Mahwah, NJ: Erlbaum.

Pfeil, U., Arjan, R., & Zaphiris, P. (2009). 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(3), 634-654.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon* 9(5), 1-6.

Quan-Haase, A. & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology and Society*, 30(5), 350-361.

Quinn, D., Chen, L., & Mulvenna, M. (2011). Does Age Make A Difference In The Behavimy Of Online Social Network Users? *Internet of Things*, IEEE, 266-272.

Ruths, D. & Pfeffer, J. (2014). Social media for large studies of behavior. *Science* 28, 346 (6213), 1063-1064.

Schroeter, R. (2012). Engaging New Digital Locals with Interactive Urban Screens to Collaboratively Improve the City. *Proceedings of the International Conference on Computer-Supported Cooperative Work and Social Computing*, ACM, 227-236.

Shao, G. (2009). Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet Research*, 19(1), 7-25.

Suh, B., Hong, L., Pirolli, P., & Chi, E. (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. *Proceedings of the International Conference on Social Computing and Networking*, IEEE, 177-184.

Sunstein, C. R. 2002. *Republic.com*. Princeton University Press.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text. *Journal of Language and Social Psychology*, 29(1), 24-54.

Thomas, M. (Ed.). 2011. *Deconstructing Digital Natives: Young people, technology, and the new literacies*. Taylor & Francis.

Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the International Conference on Web and Social Media*, AAAI.

Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior*, 9(5), 584-590.

Vieweg, S., Hughes, A.L., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazard Events: What Twitter May Contribute to Situational Awareness. *CHI*, 1079-1088.

Yardi, S. & boyd, D. (2010). Tweeting from the Town Square: Measuring Geographic Local Networks. *Proceedings of the International Conference on Web and Social Media*, AAAI.