No Reciprocity in "Liking" Photos: Analyzing *Like* Activities in Instagram

Jin Yea Jang Kyungsik Han Dongwon Lee

College of Information Sciences and Technology
The Pennsylvania State University, USA

{jzj157|kuh178|dlee}@ist.psu.edu

ABSTRACT

In social media, people often press a "Like" button to indicate their shared interest in a particular content or to acknowledge the user who posted the content. Such activities form relationships and networks among people, raising interesting questions about their unique characteristics and implications. However, little research has investigated such Likes as a main study focus. To address this lack of understanding, based on a theoretical framework, we present an analysis of the *structural*, *influential*, and *contextual* aspects of Like activities from the test datasets of 20 million users and their 2 billion Like activities in Instagram. Our study results first highlight that Like activities and networks increase exponentially, and are formed and developed by one's friends and many random users. Second, we observe that five other essential Instagram elements influence the number of Likes to different extents, but following others will not necessarily increase the number of Likes that one receives. Third, we explore the relationship between LDA-based topics and Likes, characterize two user groups—specialists and generalists—and show that specialists tend to receive more Likes and promote themselves more than generalists. We finally discuss theoretical and practical implications and future research directions.

Categories and Subject Descriptors

J.m [Computer Applications]: Miscellaneous

General Terms

Measurement, Experimentation

Keywords

Like activity, Like network, Social media analysis, Instagram

1. INTRODUCTION

The recent dramatic increase in the usage and prevalence of social media has led to the creation and sharing of a significant amount of information in various formats such as texts, photos, or videos [7][26]. It has become commonplace for people to actively access or appreciate shared content as well as to interact with the content by adding tags, comments, or Likes. A recent report by Public Broadcasting Service (PBS)[1] shows this trend in which teens and young adults are actively adding Likes and trying to receive more Likes and attention from others through Likes.

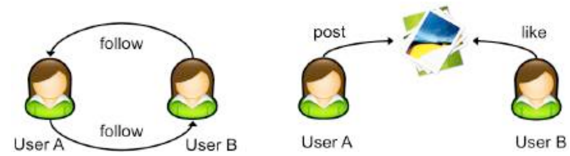

Figure 1. Two networks formed by a follow activity (left) and a Like activity (right).

In particular, Like-based interactions imply a personal preference or interest in the content shared by other users. Unlike followship-based relationships illustrated in Figure 1 (left), Liking does not necessarily require a pre-existing relationship. Instead, anyone can access and show interest by pressing a Like button on the media (Figure 1, right). Such expressions of preference from users for content in social media can take many diverse forms: for example, "LIKE" in Facebook, "+1" in Google+, "favorite" in Flickr, "re-pin" in Pinterest, and "heart" in WeChat. These micro-expressions are examples of a "content-based" relationship, and in this paper, we refer to this as *Like activities*.

Like activities imply many opportunities that can be understood in different contexts. For example, research has started to study Like activities with respect to social phenomena, because they can be interpreted as indications of one's shared interest in the content or the user who posted that content [8]. Like activities also imply business opportunities [26], where many companies take strategic approaches; for example, creating vivid and interactive posts or having more positive comments on the posts, to increase Liking on and people's interest in their brand posts [11].

Despite those research values and business implications of Like activities, however, most prior research studies have utilized them as a means to answer their other questions or hypotheses, such as a degree of popularity. Relatively, there is a lack of understanding on the intrinsic and extrinsic characteristics of Like activities, including their network structures, their relationships with other elements, and contextual aspects, which will provide unique and additional insights in the study of social media. Therefore, in this paper, we present an in-depth and comprehensive analysis on Likes and Like activities based on several test datasets drawn from the base dataset of 20 million users and their 2 billion Like activities in Instagram. We chose Instagram as a media platform, because it is currently one of the most popular social media sites with many users and ample Like activities therein. For example, a

HT '15, September 01 - 04, 2015, Guzelyurt, TRNC, Cyprus
© 2015 ACM. ISBN 978-1-4503-3395-5/15/09…$15.00
DOI: http://dx.doi.org/10.1145/2700171.2791043

[1] http://www.pbs.org/wgbh/pages/frontline/generation-like/

recent Pew research report indicates that Instagram is one of the most popular social media platforms for digital photo sharing and users with ages between 18 and 29 account for 43% of Instagram users [13].

To theoretically guide our research, we employed some research insights that have been applied in previous social media studies. The followings are our three research questions:

RQ1 (Structure): What are the structural characteristics of Like activities? How is the network formed by Like activities different from the one by Followships?

RQ2 (Influence): To what extent do user's other activities (photos, tags, comments, followers, and follows) influence Like activities?

RQ3 (Context): What are the contextual characteristics of Like activities that have been less studied?

In the balance of this paper, we first describe the theoretical framework for understanding Like activities and present previous studies on Like activities in social media. Next, we describe the process of data collection and details of the data used in the analysis. We then present our findings that show some insights on answering our research questions. Lastly, we conclude the paper with a discussion of the implications, limitations, and future work of our study and, more broadly, Likes in social media.

2. RELATED WORK

2.1 Theoretical motivations and guidelines

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, [17] 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, [9] 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. [23] 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, we show how a Like network (to be defined in Definition 1; Section 4.1) formed from Like activities is structured and developed as well as how it is different from a follow-based network, which has not been studied in social media research.

Second, when it comes to the influence of online social media, we are 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, [29] 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, [15] 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 [39]. Similar to these studies, we 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 [40], workspaces [10], and major incidents or disasters [41][42]. Studies have also indicated that social media reengineers the way of interactions between doctors and patients [20], provides richer local information to residents and facilitates local interactions [38], and helps teachers maintain professional ties with different educational communities as well as share resources and make connections with students [33]. Based on these studies, we 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. [22] presented content categories from Instagram photos; however, the small sample size (200 photos from 50 users) used in the analysis limits their findings. In our study, to infer its content, we decided to leverage tags, because users tend to add tags that meaningfully describe the photo content [21]. With this rationale, we have applied a probabilistic topic model-based tag analysis and measured the relationship between photo topics and Like activities.

Extracting photo information also allowed us 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, [35] 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. [18] 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. Since Instagram does not officially provide gender or age information, based on the topics identified through tags, we decided to characterize users through an entropy measurement and analyzed Like activities for two user groups, namely specialists and generalists.

2.2 Studies on Likes in social media

Like activities (or a similar type of content-based activities) do not necessarily need a pre-existing interpersonal relationship. Rather, it mostly asks for a similar or even the same interest in and reflection on content [24]. In this regard, the number of contents that users added Likes to could be an indication of the degree of connections. For example, when a user A added a total of 20 Likes and a user B added a total of 2 Likes to the photos posted by a user C (suppose C posted more than 20 photos), A may have more interests or stronger feelings toward C's photos (or perhaps to C) compared to B.

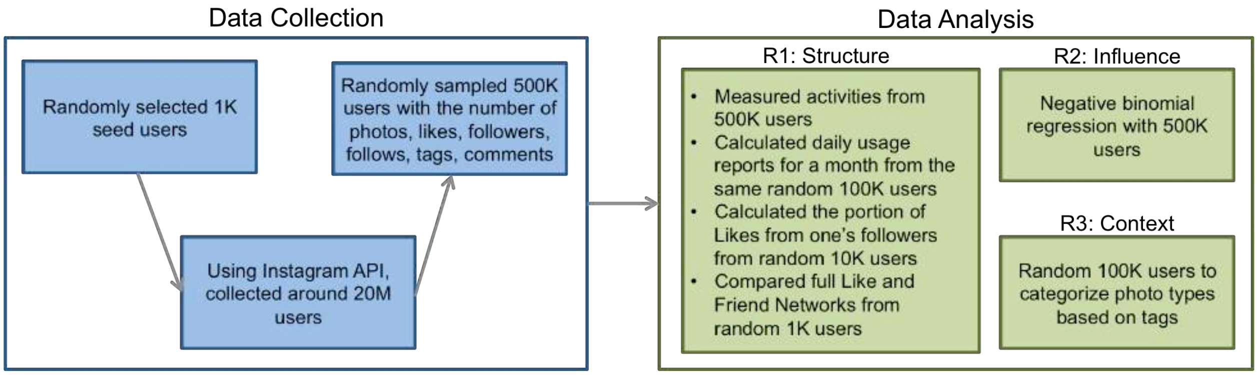

Figure 2. Workflow of data collection and analysis.

There have been research efforts on discovering the characteristics of a content-based relationship in social media. For example, [6] suggested that people tend to think that recommendations, which were derived from profile similarities and rating overlaps, are more useful and meaningful than those from one's familiarity with the recommender. [5] reported that having similar music interests can create interpersonal bonds between people. [3] studied the concept of homophily from two online social networks (BlogCatalog and Last.fm) and found that there was no significant difference among the sub-communities that were clustered based on the personal network ties for each network. This suggests that the level of one's interest is not always consistent with that of one's familiarity, and a content-based network can have more interesting items that attract people than a person-based network.

In Twitter, as retweeting has been perceived as a way of expressing one's agreement with the content, studies have shown the factors of retweeting and found that some features such as URLs and hashtags (i.e., content features), the number of followers and follows, and the age of the account (i.e., contextual features) are closely correlated to re-tweetability [8][39]. More recently, exploring the characteristics and possible applications of Like activities to predict other latent features of users in social media has appeared in [28]. They reported that Likes in Facebook could be utilized to predict user's personal traits such as gender, ethnicity, religion, etc. However, other key aspects of Likes and Like activities, such as the network structure and their relationship with other key elements in each social media site, have not been studied and investigated yet.

As its growing popularity, Instagram has gained attention from research communities, and there have been a few research studies on Like activities. Such examples include studying tag-based Like networks formed by Instagram users who have the same tags [19], studying differences in online social behaviors and engagement presented by teen and adult users in Instagram [25], exploring the relationship between the content of photos and the types of users [2], studying the relationship between online popularity of users and tag-based topical interests of their photos [14]. In every case, however, we found that Likes were primarily used as a method of measuring popularity of photos or users, where having more Likes means being more popular.

In summary, although there have been many research studies on Like activities in social media, we found that relatively little has investigated the diverse characteristics of Like activities as the main study focus. To better understand the characteristics and implications of Like activities and interactions in social media, we present our analysis from the empirical usage datasets drawn from Instagram. In particular, our rationale was to leverage existing directions used in prior studies as a theoretical guideline. We found that the application of systematic approaches toward Like activities is relatively new in social media literature; therefore, we believe that our study will provide novel insights and additional methods in the research of social media that later social media studies can leverage.

3. DATA COLLECTION

Among many social network sites, we chose Instagram for the following reasons: (1) Instagram is one of the most popular social network sites in the U.S., as reported by the Pew report [13], with a sufficient user base; (2) Compared to other social network sites such as Facebook and Twitter, Instagram has been less studied and understood; (3) With an easy interface to post photos and like others' photos, there is abundance of Like activities in Instagram; and (4) Instagram provides a well-designed and easy-to-access programming API that facilitates our data collection process.

Starting from 1,000 random seed individuals in Instagram, from March to May in 2014, we crawled to obtain the base dataset of about 20 million related users and their 2 billion Likes. To capture relatively complete Like activities across users in a chain, we exploited users' Like relationships. In other words, the dataset include those users who "liked" the photos posted by other users. In this sense, the unique aspect of our dataset is that they are all centered on users' Like activities. To collect complete Like activities, we checked each photo that a user has posted and obtained the number of Likes associated with the photo.

From user's account in Instagram, we obtained user information (e.g., user ID, name, homepage, etc.), photo information (e.g., Likes, comments, tags, etc.), and social relationship information (e.g., followers and follows). As a result, our base dataset consisted of around 20 million users and 2 billion Likes. To reduce biases from the data collection and speed up subsequent data analysis, then, we 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 2). For instance, to answer RQ1 (structure), we used 500K users to understand overall Like activities, but used 100K users to closely monitor their daily usage over a month. In addition, we 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 RQ3 (contexts), we used 100K random users.

By using different sizes of datasets for different measurements, we were 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 we 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.[2]

4. RESULTS

4.1 RQ1: Structure of Like Network

We first explored the basic structural characteristics of Like activities, namely a Like Network (LN), and their difference with the followship-based relationship, namely a Follow Network (FN). We formally define a LN as follows.

Definition 1 (Like Network): A Like Network (LN) is a directed graph $G=(V, E)$, where V is a set of users in a social network, and E is a set of directed relationships among users. An edge e_i: $u_j \rightarrow u_k$ indicates that a user u_j "Likes" a photo posted by a user u_k (u_j is not equal to u_k).

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 1. Descriptive statistics of the dataset (N=500K).

4.1.1 *Likes and other Instagram elements*

Table 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 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.

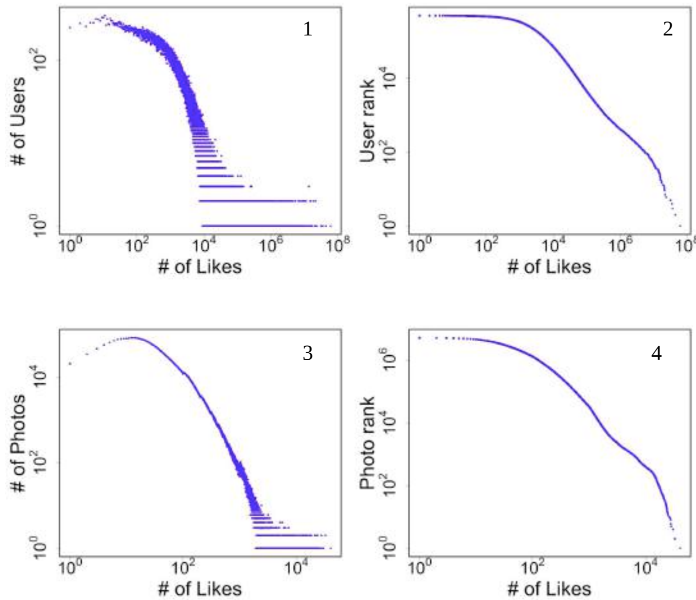

Figure 3. Distributions of # of Likes with respect to: (1) # of user, (2) user rank, (3) # of photos, and (4) photo rank. (1) and (2) are based on 500K users, and (3) and (4) are based on around 5M photos posted by randomly selected 100K users. All graphs in log-log plot.

4.1.2 *Trends of Liking and following*

We measured the trend of a LN formed over time compared to a FN. A challenge was that the Instagram API does not provide information about "when" users received Likes or started following others. To obtain this information, therefore, for the same 100K posters reported in Figure 3, we collected the total number of photos, Likes, tags, comments, followers, and follows once a day over a month period, and monitored the evolution of Like activities.

Figure 4 presents three time series graphs over a month where each data point in x-axis represents the difference (i.e., delta) of Likes, followers, and follows between two consecutive days. One finding is that all three elements increased over time in general. However, there are two differences between Likes and followers/follows. First, the average increase in Likes was much higher than the other two. For example, the number of Likes increased by 2 millions everyday on average, while the average increase in followers and follows was 93K and 15K, respectively. Second, while there was a steady increase in Likes, the variance for followers and follows somewhat fluctuated, and there was even a decrease in some days as indicated by red circles in Figure 4. This might be influenced by the way of following others in Instagram. Unlike Facebook, people do not need to be approved by others to be a friend in Instagram. Because of this, it is very

[2] Dataset is available at https://goo.gl/lyDB52

common for people to follow or un-follow others as they wish, resulting in an occasional drop in a time series.

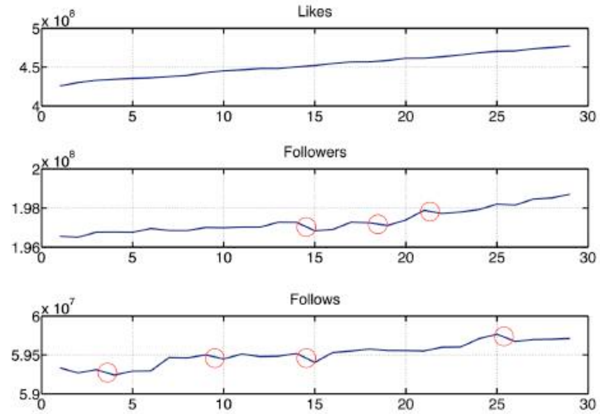

Figure 4. The difference (delta) of Likes, Followers, and Follows between two days over a month (the same 100K posters as Figure 3; x-axis is day and y-axis is daily frequency). Red circles indicate a noticeable decrease.

This occasional drop also occurs in Like activities, because "un-Liking" is also simply a one-click action that is the same as Liking. However, given that people tend to add Likes everyday whenever they like the content, but usually do not track their Like activities, they are less likely to "un-like" photos that they already "liked." In comparison, most social media sites provide an interface that allows users to manage their friends, making it easier for them to "un-follow" others. In this sense, it might be more reasonable to see a steady increase (instead of an occasional drop) in the number of Likes over time.

4.1.3 Likes from followers

Next, we investigated a LN based on the number of Likes that a poster receives from other users. From randomly chosen 10K posters, we calculated the ratio of Likes added by each poster's followers. With the IDs of users who added Likes and those of followers, we were able to check the percentage of Likes received by one's followers. As a result, interestingly, we found that almost a half of Likes were from random users with no follow-relationship (Table 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. This randomness of adding or receiving Likes implies another reason for a significant increase in a LN.

	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%)

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

Figure 5 shows the examples of the LN from two random posters, p_1 and p_2, visualized by Gephi [1], where p_1 and p_2 received a total of 10,889 and 62,821 Likes, respectively. In each figure, nodes represent other users (p_1 and p_2 not shown; located in the middle), and their sizes are proportional to the number of Likes that each node provided to the poster, p_1 or p_2. The right graph for p_2 has more users who gave many Likes to p_2, while the left graph exhibits only a few users who gave many Likes to p_1. Similar to what we have shown in Table 2, these two examples show that there are many users who only gave a "single" Like. We found that both p_1 and p_2 have received 66% and 44% of their Likes from people who added a Like only once.

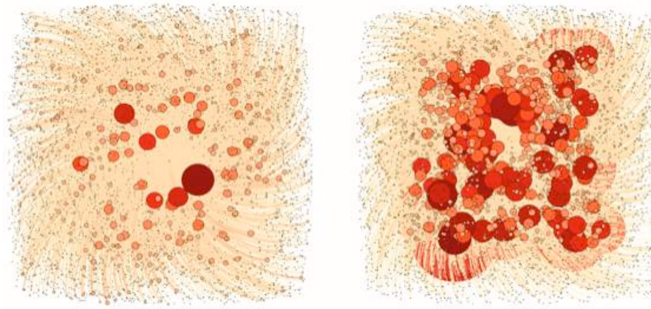

Figure 5. Example LNs of two posters, p_1 (left) and p_2 (right). p_1 and p_2 received 10,889 and 62,821 Likes, respectively. A significant number of Likes were from users who gave only a single Like (i.e., the smallest nodes) to either p_1 or p_2.

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.

4.1.4 Like and follow networks

Lastly, we compared a LN with a FN. Starting from the same 1K users, we measured the links among users up to two depths to create a FN and a LN. With these data, we again ran the network analysis using Gephi. In particular, we measured the degree of each network to show the number of links that each node has and indicate a level of network size and future growth.

	FN	LN
Avg. # Nodes	97,092	169,974
Avg. # Edges	116,444	536,599
Avg. Degree (total)	2.34	6.18
Avg. In-Degree (follower for FN; receive for LN)	1.24	3.15
Avg. Out-Degree (follow for FN; give for LN)	1.10	3.03

Table 3. Degree comparison between a FN and a LN derived from the same 1K users. All degrees were weighted.

Table 3 summarizes the degree centrality of two networks. Regarding the number of nodes and edges from the same 1K users, the LN has many more nodes and edges than the FN. There are differences in terms of degree as well. On the one hand, in-degree indicates how many followers that users have for the FN and how many Likes users received for the LN. On the other hand, out-degree indicates how many others users are following back for the FN and how many Likes that users gave back to others for the LN. As a result, both in-degree and out-degree were found to be higher in the LN than the FN, indicating that the LN tends to contain more interactions and expand more rapidly from the same number of users.

4.2 RQ2: Influences on Like Activities

Our second question examines factors that influence the number of Likes. Our 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 another perspective of Like activities. For this analysis, we 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 [2][15]. 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. We used STATA software for the analysis.

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

Table 4 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 we 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, we found that having more followers and adding more photos seem to be more influential with respect to having more Likes.

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

Table 5. LDA-discovered topics in Instagram (N=100K).

4.3 RQ3: Contexts and Like Activities

Our third question explores the contextual aspects of Like activities. In particular, we extract contextual information from photos by means of the tags in photos.

4.3.1 Topics and Likes

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, we tried to identify latent topics of the poster. To do this, we first randomly selected 100K posters. We then applied a Latent Dirichlet Allocation (LDA) model [4], using Mallet [32], an open-source machine learning toolkit, to identify a list of latent topics per poster.

As Mallet generated different topics for each execution, we 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 we found that there were some overlaps among the 100 topics, we categorized them by taking a bottom-up approach. First, to obtain ground-truth tag categories, we 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, we were 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 5 shows a list of final 20 topics with some tag examples. Lastly, to obtain the number of Likes per topic, we multiplied the ratio of each topic with the total number of Likes. We believe this was a reasonable method to find the relationship between Likes and topics, because we found that the number of Likes tends to be evenly distributed over one's photos.

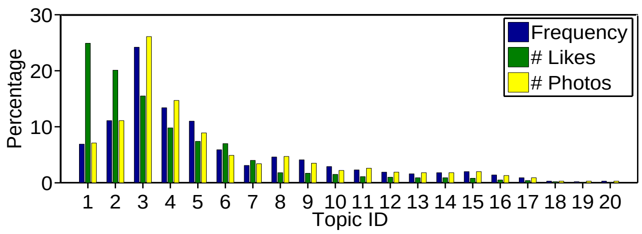

Figure 6. Ratio of frequency, # of Likes, and # of photos for the topics in Instagram (N=100K).

Figure 6 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 5) was the most frequent topic to be found in Instagram, followed by "Art/photos/design" (4th), "Fashion/beauty" (2nd), and "Holiday/vacation" (5th). We found that the tags in "Follow/shoutout/like" (10th) especially represented a unique and interesting culture in Instagram, as we 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, we 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 our 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, we believe that this result shows the connection between photo topics and Like activities in an online photo-sharing community.

4.3.2 Poster groups and Likes

We further investigated the characteristics of posters based on the topics. Once we represented each poster as a 20-dimensional topic vector via LDA, we calculated the entropy values for all 100K posters. The entropy of a poster *p* is a measure of the uncertainty in a random variable, defined as follows:

$$Entropy\ (p) = -\sum_{i=1}^{20} P(x_i)\log P(x_i)$$

where $P(x_i)$ is the probability of the topic x_i in this study. When applied to our data, a higher entropy value means that a poster *p* tends to post photos with diverse topics, while a lower entropy value means that *p* tends to post photos with a specific topic. As a result, we found that the range of entropy values is between 0 and 3.5 (see Figure 7)—that is, at minimum one topic ($2^0 = 1$) and maximally twelve topics ($2^{3.5} \approx 12$). We then defined those posters who had an entropy value smaller than 1 (i.e., less than $2^1 = 2$ topics) as specialists (those who tend to post photos with a specific theme) and those who had an entropy value higher than 3 (i.e., more than $2^3 = 8$ topics) as generalists (those who tend to post photos with diverse topics).

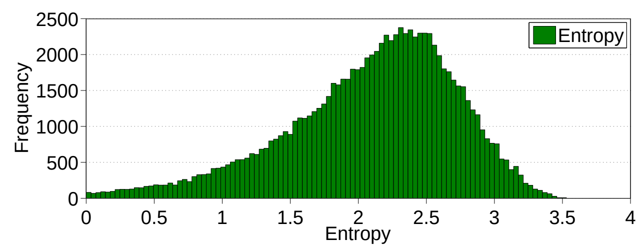

Figure 7. Distribution of entropy scores of posters (N=100K).

Table 6 shows a summary of Likes for specialists and generalists. Although the variance was high for both groups, the median results show that specialists received five times more Likes (10,893) than generalists had (2,375). In addition, the ratio of the number of Likes and the number of photos for each group indicates that on average specialists' photos appear to be more Like-inducing (e.g., high-quality landscape and architectural photos) than generalists' photos. We transformed the values to a logarithmic scale, where the t-test result showed a significant difference between two groups (i.e., $t(8823) = 133.4$, $p < 0.0001$).

Type	**# Likes**			**# Likes / # Photos (Median)**
	Mean	**Median**	**S.D.**	
Specialists (5,594)	101,666	10,893	1,358,885	35.3
Generalists (3,230)	15,989	2,375	367,911	12.5

Table 6. Summary of specialists and generalists.

Further, to verify how accurate it is to use entropy results to distinguish specialists and generalists, we randomly picked around

10% of specialists (500) and manually checked their photos based on three criteria (i.e., whether there was a specific theme in the photos; whether they have an external personal site; whether they specify a promotion site) as shown in Table 7. From the specialist group, we found that 404 posters (80.8%) have high-quality photos. However, we identified that 49 posters (9.8%; false positive) should have been considered as generalists, because we were not able to identify a specific theme from their photos. Interestingly, we found that 83 posters (16.6%) have a commercial website to advertise their photos or products (e.g., fashion items), and 129 posters (25.8%) provide a link to their personal webpage (e.g., Twitter, Facebook, Tumblr, etc.) to reach out to more people.

Question	Count (%)
How many **specialists** have a specific theme?	404 (80.8%)
How many **specialists** have a personal site?	129 (25.8%)
How many **specialists** have a promotion site?	83 (16.6%)
How many **generalists** have a specific theme?	49 (16.3%)
How many **generalists** have a personal site?	30 (10.0%)
How many **generalists** have a promotion site?	5 (1.6%)

Table 7. Differences between randomly selected specialists (N=500) and generalists (N=300) with respect to having a theme in their photos and an additional personal website.

This result implies some meaningful insights on why specialists have more Likes. It seems that a number of posters in the specialist group want to have more people visit their homepage and see their photos. Because of this, they might try to take and share high-quality photos, making their photos more distinctive, unique, and professional than others. Moreover, it seems that they tend to add tags that best describe their photos and use many tags in "Follow/shout-out/like" to get more attention (e.g., becoming a friend will expose their photos more) as well as showing their personal webpages for self-promotion or sale.

We also randomly picked around 10% of generalists (300) and checked their photos. As a result, 251 posters (83.6%) are likely to post photos that mostly describe individual stories, experiences, or thoughts. Topics in this group are diverse and more individual- and social-oriented. Such topics include "Mood/Emotion," "Social/People/Family," "Holiday/Vacation," "Sports/Activity," and "Travel." In addition, we found that photos posted by 49 posters (16.3%; false positive) have a specific theme, and thus should have been considered as specialists. There were only 5 generalists that had promotion sites (1.6%) and 30 generalists that had personal websites (10.0%). This implies that generalists tend to use Instagram to share their personal stories and experiences with others.

5. DISCUSSION

We have studied Likes and Like activities in Instagram. These small actions of adding Likes to published content (or other similar micro-activities in social media) have become a salient part of interactions in social media. As social media has become one of the primary communication channels, it has also become more natural for people to engage in "Liking." However, based on our literature review, we found that little research has studied Like activities as a main focus nor articulated their intrinsic and extrinsic aspects. In this regard, based on the research approaches applied in many social media studies, we specifically explored and studied Like activities through the lenses of their structural, influential, and contextual characteristics.

5.1 RQ1: Structure

Using large-scale Like activities in Instagram, our statistical results indicated that a LN (Like Network) expanded rapidly with respect to both the size of users and other variables (i.e., photos, comments, tags, followers, and follows). Compared to a FN (Followship Network), a LN had more nodes and links and showed a higher degree of centrality. In addition, we found that the number of Likes steadily increased over time, whereas that of followers and follows fluctuated and sometimes decreased. We also identified that a great number of Likes were from random users who only gave a "single" Like, which also accounts for the typical characteristics of a LN. We speculate that the design of Instagram (e.g., its functionality and UI; users can access not only photos by others whom they follow, but a set of random or popular photos by tags or through search) or many photo-sharing events hosted by the official Instagram blog (blog.instagram.com) might influence the size and randomness of a LN.

Another interesting implication from the randomness of Likes is that certain information can be spread very quickly and widely to a great number of people. However, most existing social media research has studied the possibility of spreading or propagating information based on one's social (followship-based) connections or networks, profile characteristics, text-based content, and so on [30][37]. In this sense, we believe that our study results open up a possibility of studying the speed and effectiveness of leveraging Like activities in spreading information to wider audience.

In addition, the structural findings that we observed here may not be repeated among other Like activities in other platforms such as Flickr, whose user demographics may be different from those of Instagram (e.g., Flickr users are generally older than Instagram users, and many professional photographers in Flickr, etc.). We leave the investigation on Like activities in other platforms for future work.

5.2 RQ2: Influence

We investigated the extent to which other Instagram factors influence Like activities. The statistical results indicated that the number of Likes was positively related to that of photos, comments, tags, and followers to different extents. In particular, followers showed the highest influence on having Likes, followed by photos, comments, and tags. Gaining attention from others was one of the motivations for using and engaging in social media [27][31]. As Public Broadcasting Service (PBS) recently noted the popularity of adding Likes among teens and young adults, and given that those populations are the main users of Instagram, our results raise an interesting question about how users tend to be involved in Like activities (e.g., the patterns of the activities; they might try to have more followers and add Likes to others' photos, etc.) compared to other older populations in social media.

Conversely, having more follows (i.e., number of users that I follow) did not show a positive influence on receiving Likes. Note that this may not be applied to every single user. However in general and when we consider a large number of users together, we can see that, even if a user decides to "follow" other users more, that action does not necessarily guarantee that the user would receive a more number of Likes from others. Perhaps this is also influenced by the fact that some very popular users have many followers and receive numerous Likes everyday, but had only few or no follows. More follow-up research is needed to find out the reasons behind the relationship between the number of Likes and the number of follows that one has

Prior studies on Twitter showed that tags and URLs exhibited the highest influence on having retweets followed by the number of followers and followings [39], which is somewhat different from our results. However, given that retweeting has many implications (e.g., Like, conversations, self-promotion, bookmark, and thanks) [16] and Liking pertains mostly to one's simple appreciations to or interests in photos, having more followers and more photos, which will increase the visibility of photos, might show a stronger effect on having more Likes.

5.3 RQ3: Context

Using topic models, based on the tags annotated to their photos, we semi-manually identified 20 prevalent topics in Instagram and found that the frequency and the number of Likes and photo for each topic were positively correlated. Based on their relationships, topics such as "Nature," "Art/photos/design," "Fashion/beauty," "Location/places/area," and "Holiday/vacation" were found to be the main ones posted, shared, and appreciated by users, which compliment the topic results from previous studies [22].

We identified two user groups, specialists and generalists, based on entropy scores for the topics. By and large, we found that specialists tend to receive more Likes in total and per photo than generalists. From the manual inspection of the samples, we found that more specialists have a personal webpage visible to others for self-promotion or sale than generalists.

Regarding topic relevance through our manual verification, our results showed around 80-85% accuracy for each group. This highlights a reasonable method of utilizing tags as a way of topic identification. It also suggests an additional feedback mechanism to Instagram users to encourage them be more engaged in online social activities. For example, Instagram can recommend a user who posts similar photos and shows similar Like activities and patterns, and users might find Like-based recommendations useful and meaningful. This design idea could further lead to creating and fostering social relationships by accessing new photo updates that are based on one's interests and adding Likes or comments.

5.4 Limitations and future work

Despite the presented interesting findings, we acknowledge a few limitations in our study. First, our results may not represent overall Like activities in Instagram or other online social media that have the Like function or a similar interface. In addition, despite the large 20 million Instagram users in the base dataset that we collected, there may exist a possibility of biases induced by the random sampling-based data collection procedure of this paper. Furthermore, by expanding our study on Like activities to other related photo-sharing social media platforms (e.g., Flickr, Pinterest, Snapchat, etc.), we want to repeat and validate if our findings are consistent across platforms. Research has also found that people show different usage patterns in social media based on the number of social media sites that they use [36], and this idea can be also applied to our study. We leave this as future work.

Second, the quantitative and statistical analysis that we have conducted cannot reveal users' motivations to add Likes and expectations to receive Likes from others, which is also a critical perspective to better understand a LN. Prior research in an online photo-sharing community [34] shows that people participate in photo-sharing activities because of intrinsic and extrinsic motivations. As our analysis already showed a relatively high correlation between comments and Likes, perhaps the texts used in comments could provide some insights on motivations for adding Likes. We are also interested in applying a qualitative approach to this idea.

Lastly, we are also interested in a co-likeness relationship, which is formed when two users liked the same photo or a user liked two photos concurrently. A similar research idea has been explored in document term co-occurrence analysis or bibliometric co-author analysis. This problem is also closely related to the collaborative filtering techniques in recommender systems. As the practical implications of such co-liked items in social media are especially high, we plan to conduct a comprehensive co-likeness analysis based on Like activities.

6. CONCLUSION

This paper contributed to an exploration and articulation of one of the most popular activities in social media—"Liking"—according to three research perspectives—structure, influence, and context. Using several datasets of different sizes that were randomly drawn from a base of 20 million users and 2 billion Likes in Instagram, we found that a Like network (LN) is a fast-expanding network, formed and developed by both one's friends and random users. We found that five other Instagram elements influence the number of Likes received to different extents. In addition, using an LDA-based tag analysis, we identified 20 latent topics, prevalent among tags added to photos, and presented top 5 topics in Instagram. Furthermore, we distinguished among posters with special topics (specialists) and those with diverse topics (generalists) and found that specialists tend to receive more Likes and provide additional channels for self-promotion, whereas generalists showed opposite characteristics.

7. ACKNOWLEDGEMENT

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

8. REFERENCES

[1] Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 361-362.

[2] Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on instagram. *CHI*, 965-974.

[3] Bisgin, H., Agarwal, N., & Xu, X. (2010). Investigating homophily in online social networks. *Web Intelligence and Intelligent Agent Technology*, 533-536.

[4] Blei, D.M, Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *J. of Machine Learning Research*, 3, 993-1022.

[5] Boer, D., Fischer, R., Strack, M., Bond, M., Lo, E., & Lam, J. (2011). How Shared Preferences in Music Create Bonds Between People: Values as the Missing Link. *Personality and Social Psychology Bulletin*, 1-13.

[6] Bonhard, P., Harries, C., McCarthy. J., & Sasse, A. (2006). Accounting for Taste: Using Profile Similarity to Improve Recommender Systems. *CHI*, 1057-1066.

[7] boyd, d. & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *J. of Computer-Mediated Communication*, 13(1), 210-230.

[8] boyd, d., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *HICSS*, 1-10.

[9] Chen Y., Chuang, C., & Chiu, Y. (2014). Community detection based on social interactions in a social network. *JASIST*, 65(3), 539-550.

[10] 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. *JASIST*, 65(10), 2035-2047.

[11] De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: an investigation of the effects of social media marketing. *J. of Interactive Marketing*, 26(2), 83-91.

[12] Duggan, M. & Brenner, J. (2012). The Demographics of Social Media Users - 2012. *Pew Research Center's Internet & American Life Project*.

[13] Duggan, M. (2013). Photo and Video Sharing Grow Online. *Pew Research Center's Internet & American Life Project*.

[14] Ferrara, E., Interdonato, R., & Tagarelli, A. (2014). Online popularity and topical interests through the lens of instagram. *HT*, 24-34.

[15] Gilbert, E., Bakhshi, S., Chang, S., & Terveen, L. (2013). "I Need to Try This!": A Statistical Overview of Pinterest. *CHI*, 2427-2436.

[16] Gorrell, G. & Bontcheva, K. (2015). Classifying Twitter favorites: Like, bookmark, or Thanks?. *JASIST*.

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

[18] 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.

[19] Han, K., Jang, J., & Lee, D. (2015). Exploring Tag-based Like Networks. *CHI*, 1941-1946.

[20] 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.

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

[22] Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we instagram: A first analysis of instagram photo content and user types. *ICWSM*.

[23] 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).

[24] Jacovi, M. et al. (2011). Digital Traces of Interest: Deriving Interest Relationships from Social Media Interactions. *ECSCW*, 21-40.

[25] Jang, J., Han, K., Shih, P. C., & Lee, D. (2015). Generation Like: Comparative Characteristics in Instagram. *CHI*, 4039-4042.

[26] Kaplan, A.M. & Haenlein, M. (2009). Users of the world, unite! The challenges and opportunities of Social Media. *J. of Business Horizons*, 53(1), 59-68.

[27] Kietzmann, J., Hermkens, K., McCarthy, I., & Silvestre, B. (2011) Social media? Get serious! Understanding the functional building blocks of social media. *J. Business Horizons*, 54, 241-251.

[28] Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 110(15), 5733-5734.

[29] Lampe, C., Ellison, N., & Steinfield, C. (2007). A Familiar Face(book): Profile Elements as Signals in an Online Social Network. *CHI*, 435-444.

[30] Lee, K., Mahmud, J., Chen, J., Zhou, M., & Nichols, J. (2014). Who Will Retweet This? Automatically Identifying and Engaging Strangers on Twitter to Spread Information. *IUI*, 247-256.

[31] Leung, L. (2009). User-generated content on the internet: an examination of gratifications, civic engagement and psychological empowerment. *New Media & Society*, 11(8), 1327-1347.

[32] McCallum, A. K. (2002). "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu/

[33] 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*.

[34] Nov, O., Naanam, M., & Ye, C. (2009). Analysis of Participation in an Online Photo-Sharing Community: A Multidimensional Perspective. *JASIST*, 61(3), 555-566.

[35] Ottoni, R., Pesce, J., Casas, D., Franciscani, G., Meira, W., Kumaraguru, P., & Almeida, V. (2013). Ladies First: Analyzing Gender Roles and Behaviors in Pinterst. *ICWSM*.

[36] Petrocchi, N., Asnaani, A., Martinez, A., Nadikarni, A., & Hofmann, S. (2015). Differences between People who Use Only Facebook and Those who Use Facebook Plus Twitter. *J. of Human-Computer Interaction*, 31(2), 157-165.

[37] Pfitzner, R., Garas, A., & Schweitzer, F. (2012). Emotional Divergence Influences Information Spreading in Twitter. *ICWSM*.

[38] Schroeter, R. (2012). Engaging New Digital Locals with Interactive Urban Screens to Collaboratively Improve the City. *CSCW*, 227-236.

[39] Suh, B., Hong, L., Pirolli, P., & Chi, E. (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. *SocialCom*, 177-184.

[40] Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*.

[41] 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.

[42] Yardi, S. & boyd, D. (2010). Tweeting from the Town Square: Measuring Geographic Local Networks. *ICWSM*.