

Does #like4like Indeed Provoke More Likes?

Yang Zhang
CISPA, Saarland University
Saarland Informatics Campus

Minyue Ni
Software School
Shanghai Key Laboratory of Data Science
Fudan University

Weili Han
Software School
Shanghai Key Laboratory of Data Science
Fudan University

Jun Pang
FSTC & SnT
University of Luxembourg

ABSTRACT

Hashtags, created by social network users, have gained a huge popularity in recent years. As a kind of metatag for organizing information, hashtags in online social networks, especially in Instagram, have greatly facilitated users' interactions. In recent years, academia starts to use hashtags to reshape our understandings on how users interact with each other. #like4like is one of the most popular hashtags in Instagram with more than 290 million photos appended with it, when a publisher uses #like4like in one photo, it means that he will like back photos of those who like this photo. Different from other hashtags, #like4like implies an interaction between a photo's publisher and a user who likes this photo, and both of them aim to attract likes in Instagram. In this paper, we study whether #like4like indeed serves the purpose it is created for, i.e., *will #like4like provoke more likes?* We first perform a general analysis of #like4like with 1.8 million photos collected from Instagram, and discover that its quantity has dramatically increased by 1,300 times from 2012 to 2016. Then, we study whether #like4like will attract likes for photo publishers; results show that it is not #like4like but actually photo contents attract more likes, and the lifespan of a #like4like photo is quite limited. In the end, we study whether users who like #like4like photos will receive likes from #like4like publishers. However, results show that more than 90% of the publishers do not keep their promises, i.e., they will not like back others who like their #like4like photos; and for those who keep their promises, the photos which they like back are often randomly selected.

CCS CONCEPTS

• **Information systems** → **Social networks**; • **Applied computing** → *Sociology*;

KEYWORDS

Social networks; data mining; hashtags; social media behavior

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

WI '17, August 23-26, 2017, Leipzig, Germany
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-4951-2/17/08...\$15.00
<https://doi.org/10.1145/3106426.3106460>

ACM Reference format:

Yang Zhang, Minyue Ni, Weili Han, and Jun Pang. 2017. Does #like4like Indeed Provoke More Likes?. In *Proceedings of WI '17, Leipzig, Germany, August 23-26, 2017*, 8 pages.
<https://doi.org/10.1145/3106426.3106460>

1 INTRODUCTION

Online social networks (OSNs) have been one of the most successful applications during the past decade. Leading social network companies have gained a large number of users. To present some statistics, Facebook has 1.65 billion monthly active users; 200 billion tweets are shared by Twitter users every year; and more than 95 million photos are published in Instagram on a daily base. OSNs are changing the way of our life in many perspectives, including how we share life moments, maintaining friendship and communicating with others. In addition to that, people have also created many new concepts in social network platforms, and one such example is *hashtag*. In simple terms, a hashtag is a word or unspaced phrase preceded by the symbol #, it was created by Chris Messina, an open web advocate, back in 2007 in Twitter, and quickly adopted by users from all social networks. Nowadays, almost all OSNs provide users with the hashtag function. Hashtag is proposed to serve as a metatag for organizing information, in order to allow users to efficiently find data in OSNs with a specific theme. With its becoming more and more popular, a hashtag's function has expanded well beyond that. Nowadays, many people use hashtags to organize events. For instance, during the Paris terrorist attack in 2015, people living in Paris use #porteouverte in OSNs to offer safety shelters to people on the street; the international activist movement campaigning against systemic racism and violence towards African Americans started from the use of #blacklivesmatter.

Among all the leading OSNs, Instagram users concentrate the most on hashtags. It is very common for an Instagram photo to be appended with around 20 different hashtags, such as the photo in Figure 1. Moreover, Instagram users have invented many new hashtags as a way of social interactions. Popular hashtags of this kind include #instagood, #tbt, #followme, #nofilter. The amount of hashtags is increasing exponentially in Instagram (there are more than one billion photos in Instagram appended with the above four hashtags), and these hashtags have led to a great opportunity to understand users' behaviors in OSNs. A recent work [16] has used #selfie to study the phenomenal self-portrait convention in Instagram, while the authors of [9] have analyzed users' dining behaviors worldwide with #foodporn.

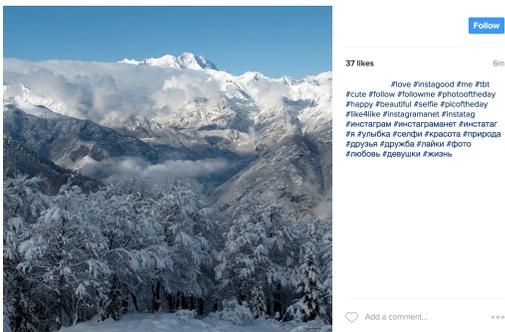


Figure 1: An Instagram photo.

Among the popular hashtags, one hashtag particularly draws our attention, namely, #like4like. “Like” is an action that users in OSNs can perform when they see certain posts that are interesting to them. In many cases, the popularity of certain posts or users in OSNs are evaluated by simply counting the number of likes they receive. The hashtag #like4like is the abbreviation for “like for like”: if a user publishes a photo (*publisher*) appended with #like4like, then he will (or promises to) like back photos of others who like this photo (*likers*). #like4like appears with the birth of Instagram, and it is one of the most popular hashtags at the moment. By now, there exist more than 290 million photos with #like4like as one of their hashtags in Instagram.

The semantics of #like4like implies an interaction between publishers and likers, both of which share the same purpose of attracting more likes. Also, the interaction is sequential, i.e., a liker can only receives *like-backs* from a publisher if he (or she) likes the publisher’s photo first. The dynamic interaction distinguishes #like4like from other popular hashtags in Instagram, such as #love, #selfie and #instagood, and drives our curiosity: *will #like4like really provoke more likes for both publishers and likers?* Since both publishers and likers aim to get more likes through #like4like and there is a causal relation between the behaviors “like” and “like back”, we decompose our main objective into two research questions:

RQ1. Will a publisher get more likes when publishing a photo appended with #like4like?

RQ2. Will a liker receive like-backs from a publisher if he liked the publisher’s #like4like photo?

The current work. By analyzing a large dataset of #like4like photos (1.8 million) collected from Instagram, we perform a general analysis on the rise of #like4like. Our observations include: the quantity of #like4like has increased by 1,300 times from 2012 to 2016, meanwhile the number of users who use #like4like has increased 600 times; female users take a larger proportion for using #like4like in the beginning, but male users are catching up.

To answer the first research question (*RQ1*), we perform four analyses. First, we discover that in the early stage of Instagram (from 2012 to 2013), #like4like indeed attracted more likes for their publishers than other hashtags. However, since 2014 the data shows that photos with and without #like4like hashtag exhibit no big difference in attracting new likes. Second, among all the photos

appended with #like4like, those concentrating on health and traveling themes (summarized from other hashtags) get more likes than other photo types. This indicates that the content of a photo is the most important factor to attract likes, not #like4like itself. Third, we observe that users who like #like4like photos are more similar to the publishers, especially in terms of gender homophily. Fourth, by analyzing the dynamic interaction between publishers and likers, we discover that the the lifespan of #like4like photos is quite limited: 55.32% of the likes they receive are within one hour after being published.

To answer the second research question (*RQ2*), we perform two analyses. First, we discover that most of the #like4like publishers do not keep their promises (by liking back any photos of their likers), only 6.08% total likes have received like-backs. And those likers who actually receive the likes from a publisher are more likely to be those who used to have interactions with the publisher. Second, by analyzing publishers’ liking back behaviors, we discover that most publishers simply like the first photo of the likers, and contents of the likers’ photos are rather random.

Paper structure. In the rest of the paper, Section 2 introduces the dataset we have collected for conducting analysis. Then, we perform a general analysis on the rise of #like4like in Instagram in Section 3. Section 4 addresses the first research question on whether #like4like attracts more likes for their publishers. Section 5 studies the second research question on whether publishers will keep their promises to like their likers’ photos back. Section 6 summarizes some related works and we conclude the paper in Section 7.

2 DATASET

We collect our dataset from Instagram through its REST API¹ on April, 2016. First, to obtain an unbiased Instagram user sample, we adopt a strategy similar to [16], i.e., randomly sampling user IDs. We concentrate on sampling numbers below 1.6 billion since the work [16] points out that no Instagram users’ IDs are above 1.6 billion. For each randomly sampled number, we use the function “/users/user-id” of Instagram’s API to check whether it is a valid Instagram user ID or not, in the end we obtain 11,982,242 users. Then, for all these 11 million users, we use Instagram’s API to extract all of their published photos’ information including publishing time, number of likes, hashtags, etc. Meanwhile, we also collect these users’ profile information including their number of photos, number of follower/followees, and profile photos. In the end, we collect 424,339,455 photos. As one important aspect of our analysis is the location information of photos with #like4like, thus we also collect the geographical coordinates (longitude and latitude) of each photo if available, in total we get 40,967,395 photos with location information. We further adopt the API of Foursquare², a popular location-based social network service, to map each geographical coordinates to a country.

Among all the 40 million photos collected, we obtain 1,822,225 photos appended with either #like4like or #l4l (a short version of #like4like) from 143,586 users for our analysis. It is worth noticing that the authors of [16] concentrate on analyzing #selfie in Instagram and they have adopted several methods to construct their

¹<https://www.instagram.com/developer/>

²<https://developer.foursquare.com/>

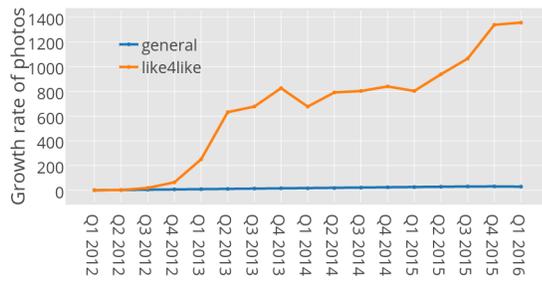


Figure 2: The growth rate of numbers of Instagram photos (2012 Q1-2016 Q1).

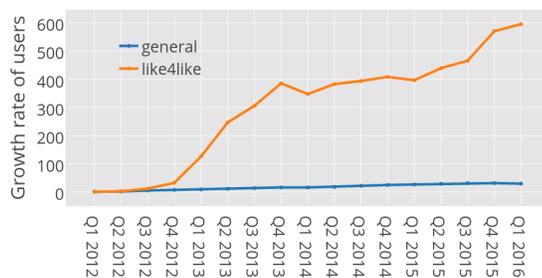


Figure 3: The growth rate of publishers (2012 Q1-2016 Q1).

dataset besides directly finding photos appended with #selfie, such as selfie in different languages, hashtags that are similar to selfie, e.g., #me, and photos with human faces. However, none of these methods are suitable in our work: as an Instagram created notion, #like4like is younger than #selfie and we have not found its corresponding versions in other languages; similarly, there exist no hashtags that are semantically close to #like4like; different from #selfie, #like4like does not regulate contents of the photo being tagged, thus we do not need to analyze photos to find #like4like.

To obtain users' demographics, we use Face++'s API to recognize the profile photo of each user who has shared #like4like to get his gender, race (Asian, African and White) and age. Face++ is a state-of-the-art facial recognition service based on deep learning techniques, it has been adopted in many works to obtain users' demographic information such as [10, 11, 15, 16].

As discussed before, one distinct property of #like4like, compared to other hashtags, is that #like4like implies a dynamic interaction between publishers and likers. Therefore, to fully understand #like4like, we cannot neglect this interaction. Unfortunately, Instagram's API does not provide the time information when a photo is liked, thus we adopt a different approach to collect the data. We first use the "/tags/tag-name/media/recent" function of Instagram's API to get a randomly chosen photo appended with #like4like published by a user (e.g., Alice) recently, then we keep on "monitoring" the photo to see when it gets new likes from other users. Once we observe a new like, e.g., from a user Bob, we record the time as the time Bob likes Alice's photo. Next, besides monitoring the photo of Alice for obtaining other users' likes, we also start to monitor Bob's

latest 30³ photos he published in Instagram. If one of Bob's photos gets a new like from Alice, then we record the time as the time that Alice likes back. Note that to prevent the case that Alice liked Bob's photos before, prior to monitoring Bob's photos, we first search all the likes of Bob's recent 30 photos, if Alice liked any of these photos, we record it in a separate file as "precursory likes". Later in this paper (Section 4), we will use "precursory likes" as a kind of weak interaction between Alice and Bob for further analysis. We pick multiple random time to start our crawling script during the first two weeks of April, 2016, each script is executed for 3 days. In the end, we get 331 photos appended with #like4like from 331 different users and these photos in total receive 19,086 likes.

For presentation purposes, we refer the 331 #like4like photos with 'like' and 'like back' information as the dynamic dataset and the 1,822,225 #like4like photos collected from randomly sampling users as the static dataset. Both datasets are available upon request.

3 THE RISE OF #LIKE4LIKE IN INSTAGRAM

We first study the rise of #like4like photos in the temporal space with the static dataset. Then, we concentrate on #like4like publishers, with respect to their demographics and user groups.

3.1 How fast does #like4like grow?

As discussed in Section 1, #like4like is one of the most popular hashtags in Instagram. We are interested in its growth pattern during the past few years. We calculate #like4like photos' growth rate from the first quarter in 2012 (Q1 2012) to the first quarter in 2016 (Q1 2016), where the number of #like4like photos published in Q1 2012 is used as the base: the growth rate in Q1 2012 is indicated as one and higher the growth rate, the faster the growth of #like4like. To validate our results, we calculate the growth rate of a same size random photo sample obtained from the general Instagram photos (424,339,455), the results are presented in Figure 2. As we can see, the number of #like4like photos has increased significantly during the past few years, with the number of #like4like photos in Q1 2016 being 1,300 times more than it was in Q1 2012. Especially, we observe a dramatic increase in the first half of 2013. Meanwhile, with Instagram being more and more popular, the number of photos published in Instagram is also increasing rapidly (about 30 times increased from 2012 to 2016); but when compared to #like4like, this growth rate can be neglected (see the bottom blue line in Figure 2). We also calculate the growth rate of users for #like4like and use the growth rate of users from the same random sample as the baseline, and the results are described in Figure 3. The increase is still very fast (600 times), even though not as significant as in Figure 2. The difference growth rate between #like4like photos and #like4like publishers also shows that many users publish #like4like photos multiple times. In summary, #like4like appears with the birth of Instagram and within a few years it has become significantly popular.

3.2 Who publish #like4like photos?

Demographics. Figure 4 depicts the distribution of #like4like publishers according to their gender, race and age. Since we did not collect the demographics of general users, our analysis concentrates

³30 is the maximal number of photos Instagram provides on a single query.

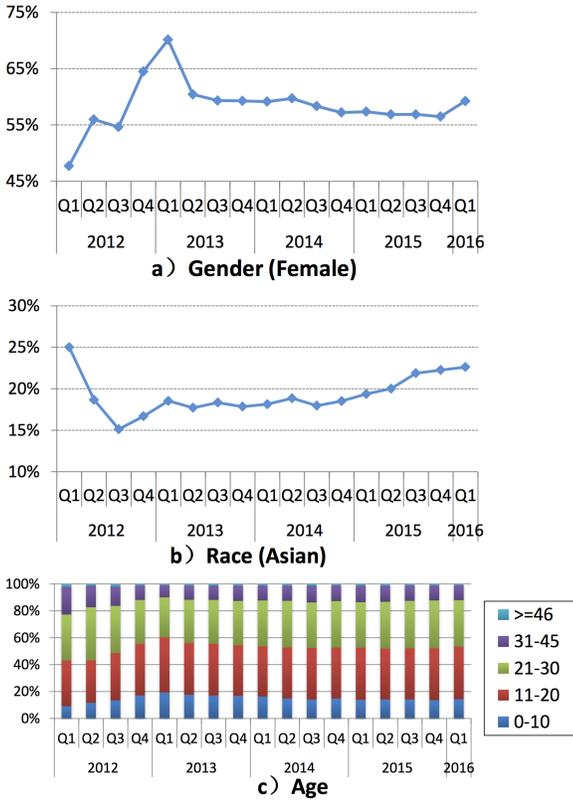


Figure 4: Distribution of users' demographics.

on the trend within each demographic group. From Figure 4, we can see that the proportion of female users publishing #like4like is decreasing, meaning that #like4like is gaining popularity among male users. With respect to race, we find that the percentage of Asian users is increasing from 15.12% to 22.65% by the first quarter in 2016. Accordingly, the percentage of white users is decreasing, while that of African users stay relatively stable. We observe from Figure 4 that more and more young users start to use #like4like. The percentage of users aged under 20 increases from 43% to 53%, while for users aged between 11 and 20 it is about 39%. On the contrary, for users aged above 30 it is decreasing from 22% to 12%.

User types. Altogether, there are 143,586 users who have published #like4like photos in our static dataset. Following the same methodology of [9], we divide them into three groups:

- Singletons: those who have published #like4like photos only once.
- Travelers: those who have published #like4like photos in multiple countries.
- General users: this includes users who have ever published #like4like photos.

Table 1 lists the basic statistics about users in different groups. We can see that although there are only 2,635 travelers (1.84%), the total #like4like photos they publish is 137,533, which is about 7.77% of the total number of #like4like photos. Meanwhile, singletons take a much large proportion (about 19 times more than travelers)

Table 1: Statistics of different user groups.

	#.(%) users	#.(%) photos	avg. #. photos
Singletons	50,311 (35.04%)	50,311 (2.76%)	1
Travellers	2,635 (1.84%)	137,533 (7.54%)	52.19
General	143,586	1,822,225	12.33

in #like4like publishers, but their #like4like photos only take less than 1/3 of that of travelers. In addition, we observe that travelers' average number of #like4like photos is much higher than that of the general one (12.33). The results indicate that people who have traveled in multiple countries are more willing to publish photos appended with #like4like.

4 RQ1: WILL #LIKE4LIKE ATTRACT LIKES FOR PUBLISHERS?

In this section, we aim to answer the first research question – *will #like4like attract more likes for publishers?* Towards this goal, we perform four analyses. First, we study the number of likes a #like4like photo gets; second, we check what kind of #like4like photos will attract more likes with respect to their contents and publishers; third, we concentrate on who will like #like4like photos; fourth, we study the lifespan of a #like4like photo in terms of the likes it gets in the temporal dimension.

4.1 How many likes attracted by #like4like?

As discussed in Section 1, the purpose of a #like4like photo publisher is to attract more likes for himself. We want to know whether the hashtag #like4like indeed serves this purpose. Through analysis on the static dataset, we observe that the average number of likes a #like4like photo receives is 73.99, while for a general Instagram photo, this number is 46.06, i.e., #like4like gains 62% more likes for its publishers. This seems to suggest that to get more likes, publishing a #like4like photo is a good choice. However, we will draw a different conclusion if we check the average number of likes over time. As shown in Figure 5, when Instagram is in its early stage, i.e., from Q1 2012 to Q1 2013, #like4like photos indeed serve their purposes by getting 3 times more likes for their publishers than normal photos (the same random sample served as baselines in Section 3.1). However, after that, the difference started to disappear, i.e., photos with or without #like4like, have no difference in terms of the number of likes they receive. This indicates that #like4like nowadays will not get extra likes for users. We believe that this is due to the maturity of Instagram users who, after using Instagram for a while, do not need to accumulate likes to establish their social status anymore. Nevertheless, #like4like is still a popular hashtag (see Figure 2), and the number of users who use it is still increasing (see Figure 3). In the next section we investigate what makes a successful #like4like photo in terms of the number of likes it attracts.

4.2 What actually attracts likes?

Through two aspects, we study what makes a #like4like photo get more likes, including photo contents and their publishers.

#like4like photo contents. To obtain the contents of photos, manual annotation is necessary but requires a lot of time and resources. On the other hand, the authors of [4] recently show that more than

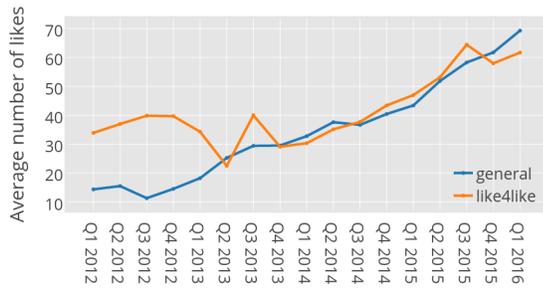


Figure 5: Average number of likes (2012 Q1-2016 Q1).

50% of the Instagram hashtags are related to the contents of photos they append to. Therefore, in our study, we treat other hashtags in #like4like photos as their annotation.

Following the results of [9]⁴, we partition all hashtags into seven categories including *Sentiment*, *Health*, *Social*, *Location*, *Food*, *Time-date* and *Others*. Table 2 lists the amount and some examples of each category. Then, with these defined hashtag categories, we organize photos into seven categories as well, where a photo is classified into one category as long as it is appended with a hashtag belonging to that category. For a photo appended with hashtags of multiple categories, we count it into all the categories its hashtags belonging to. Figure 6 and Figure 7 show the average numbers of likes and comments received by #like4like photos in each category. From Figure 6, we discover that, health photos have received the most likes (52.51 likes in average, 36 in median), followed by location photos. Meanwhile, photos related to food receive the least number of likes. For the average number of comments received by photos of different categories, there does not exist a big difference. This is due to commenting on other users' photos is inherently interactive, thus requires users' creativity which many users do not bother to do. Nevertheless, health photos still gain more comments than photos of other categories.

Table 2: The seven categories for hashtags.

Category	Amount	Examples
Emotions	115	like4like, l4l, happy, love
Health	136	nutrition, organic, sport
Social	134	friend, igers, holiday, instagood
Locations	810	jarkarta, italy, japanese, home
Foods	657	banana, beef, beer, breakfast
Date & Time	52	summer, monday, midnight
Others	176	adventure, heaven, homemade

#like4like publishers. Next, we focus on which kind of users get more likes when publishing #like4like photos. We calculate the average number of likes #like4like photos get with respect to the previous presented three types of users, i.e., singletons, travelers and general users. As shown in Table 3, travelers' #like4like photos receive more likes than singletons (53.07 vs. 35.01) and the number is also much higher than the general user group. This suggests that travelers do not only publish more #like4like photos, but also

⁴The detailed hashtag category can be downloaded at <https://tinyurl.com/foodporn-hashtags>.

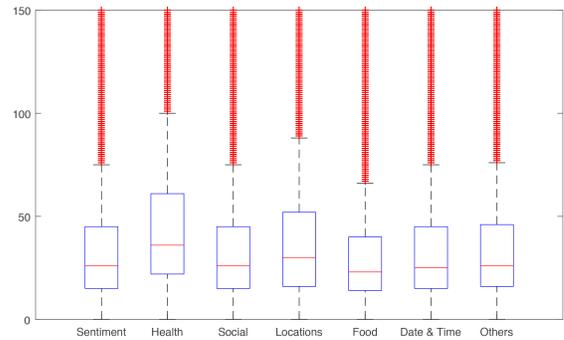


Figure 6: Average numbers of likes for each category.

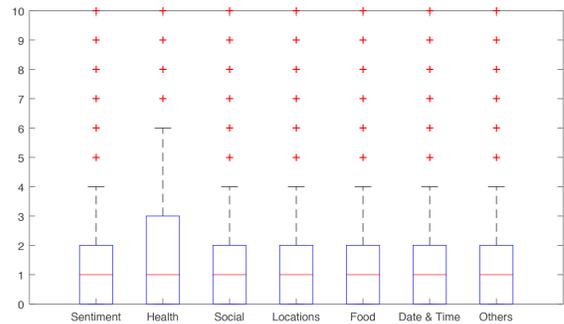


Figure 7: Average numbers of comments for each category.

get more likes through #like4like. To understand this in detail, we compare the contents of #like4like photos published by travelers and singletons. Table 4 shows that travelers tend to use more location photos than singletons, which is reasonable since they have checked in multiple countries. As location photos receive the second most likes, this can explain why travelers get more likes than singletons. Meanwhile, we also see that singletons use more sentiment photos than travelers, as shown in Figure 6 sentiment photos are among the least favorable #like4like photos for receiving likes, thus singletons' #like4like is not very attractive.

Table 3: Detailed posts information of users of each type.

	avg. #. likes	avg. #. comments
Singleton	35.01	1.87
Travelers	53.07	1.94
General	39.29	1.61

Table 4: Distributions of photo categories used by singletons (singl.) and travelers (travel.).

	sent.	health	social	loc.	food	time	rest
singl.	41.83	0.91	24.84	2.35	3.16	2.70	24.21
travel.	37.85	0.94	25.97	3.69	3.41	2.80	25.34

In summary, #like4like photos on health and location topics are easier to get likes and travelers' #like4like photos are popular.

4.3 Who likes #like4like?

After analyzing photo contents and publishers, we study users who like #like4like photos through demographic homophily and user social relations.

Our first hypothesis is that users tend to like #like4like photos whose publishers are similar to them. Social homophily is one of the most important hypotheses for understanding user behaviors in online social networks. It describes the tendency of individuals to interact with others who are similar to themselves. Following [16], we define gender homophily as the following

$$H_{gender} = \frac{F_{mm} + F_{ff}}{F_{interaction}}$$

where F_{ff} (F_{mm}) represents the number of likes from female (male) likers to #like4like photos published by female (male) users and $F_{interaction}$ stands for the total number of likes for #like4like photos. Higher the value of H_{gender} , stronger the gender homophily. To validate our result, we build a baseline model by randomly organizing #like4like publishers and those who have liked any #like4like photos into random pairs, and compute the corresponding homophily. The race homophily and its baseline model are defined accordingly. Since age is not categorical, we adopt the inverse of root-mean-square error (RMSE) as the measurement [16].

As shown from Figure 8, both gender and race homophily are stronger than the those of the two baseline models, which means the likers tend to like #like4like photos published by users with the same gender and race. Figure 8 further indicates the trend of age homophily as well. What is quite different from gender and race is that, the difference between H_{age} and its baseline is smaller. Statistics shows that among all users on Instagram, about 90% of whom are aged between 20 to 40. Thus the age difference between any two of the random users on Instagram is limited. So this could explain that H_{age} is not much stronger than its baseline.

Besides demographic homophily, we also examine the relation between publishers and likers. Since during the dynamic data collection, we have collected the information whether a publisher liked any of a liker's photo before, i.e., *precursory likes*, thus we use these precursory likes to define the relation between publishers and likers. Concretely, we categorize the publisher-liker relation into the following two types.

- *Friends*. If a publisher has liked a liker's photo before, then they are considered as friends.
- *Strangers*. If a publisher has never liked a liker's photos before, then they are considered strangers.

After investigation, we find that #like4like does help attract attention from strangers, 80.02% of likes are from strangers. This result is within our expectation since #like4like is a very popular hashtag in Instagram, which would lead to more strangers' attention.

4.4 How long does #like4like sustain?

When a #like4like photo is published, we want to know when it will get likes, and how long this like period will sustain. Since Instagram's API does not provide the time when a user likes a photo as addressed in Section 2, we use the dynamic dataset we collect to answer this question.

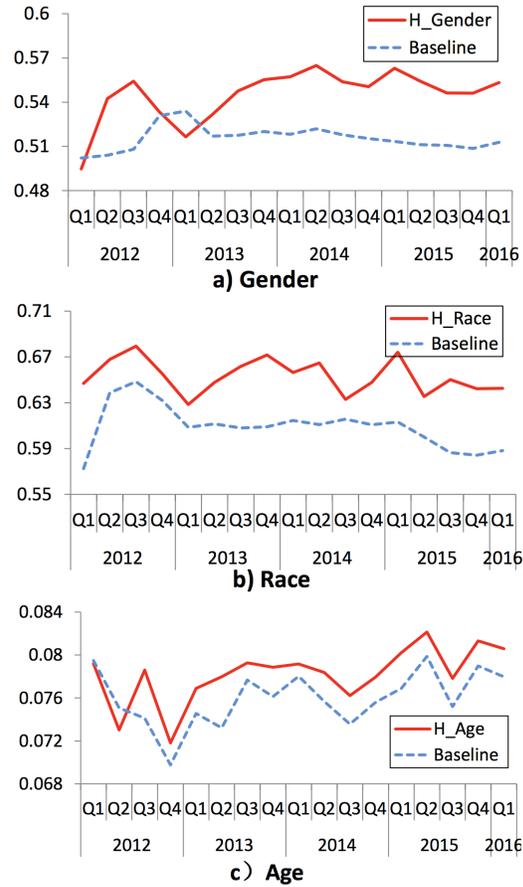


Figure 8: Demographic homophily and baselines.

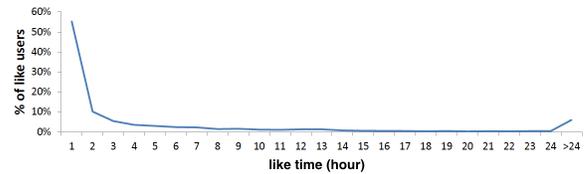


Figure 9: Time distribution to like #like4like photos.

Figure 9 shows the distribution of time difference when #like4like photos receive likes after being published. As we can see, #like4like photos receive more than half of their likes (55.32%) within one hour, and over 94% of likes within 24 hours. This indicates that #like4like is highly momentum, thus to get more likes through #like4like, a user should pick a time when Instagram users are active online.

In conclusion, we discover in this section that #like4like in the beginning will get more likes for its publishers but not many more after that; health and location related #like4like photos receive more likes than photos of other types; travelers' #like4like photos receive more likes than singletons; users tend to like #like4like photos whose publishers are similar to them with respect to demographics; the lifespan of a #like4like photo is usually quite short.

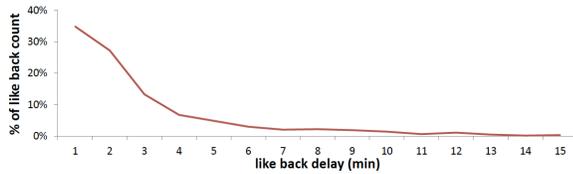


Figure 10: Time distribution for publishers to like back.

5 RQ2: WILL PUBLISHERS LIKE BACK?

Both #like4like publishers and likers aim to get more likes. In the previous section, we have concentrated on publishers, while in this section we study #like4like likers, i.e., *will a liker get likes from a publisher if he has liked the publisher's #like4like photo?*

We start by checking the proportion of likers that get like-backs, then study the time when the publishers like back likers' photos. Next, we examine the relation between publishers and likers. In the end, we study the types of photos that receive like-backs.

5.1 How many likes get like-backs?

In general, among all the 19,086 likes on the 331 #like4like photos we have collected in the dynamic dataset, a very low percentage, only 6.08% of them got #like4like publishers' like-back. For all the 331 publishers, more than 52% of them did not like back any of their likers, and 17.22% of them liked back only one liker. Meanwhile, only 10.27% of the publishers have liked back more than 5 likers. Table 5 lists the distribution of publishers who like back their likers. This result shows that #like4like publishers do not keep their promise to like back their likers. In another way, there is no obvious tendency for them to socialize and interact with users who like their photos. What they really want through #like4like is just to attract more likes for themselves. This observation may also explain why #like4like will not get more likes for publishers in Instagram after 2013.

Table 5: The distribution of publishers who like back.

	0	1	2-5	>5
# of users	175	57	66	34
%	52.87	17.22	19.94	10.27

5.2 When to like back?

For all the publishers who like back their likers, we study when their like-backs take place. As depicted in Figure 10, most of them like back the photos of likers within 15 minutes. More surprisingly, nearly 90% of like-backs happen within 5 minutes. Compared to the result in Figure 9, publishers' like-back is much faster than likers' like. This suggests that after a user publishes a #like4like photo, he is likely to stay online to receive likes and directly likes back the likers. As likes come slower, this may be another reason why most of the likers do not receive any likes.

5.3 Whom to like back?

Even though only 6.11% of the users have received like-backs, we are still interested in who these users are. We tackle this question through the relationship between likers and publishers, i.e., friends

and strangers (see Section 4). Through analysis, we discover that among the 6.11% users, 67.75% of them are friends of #like4like publishers. The result is even more significant by considering that in average, the friends of a publisher only contribute 19.58% of the likes he receives. Based on this, we can see that users would be more likely to like back their friends' photos, and it would be less possible for them to react to strangers' likes. This means that popular hashtags like #like4like cannot influence users' attitude towards strangers in social networks.

5.4 What to like back?

Finally, we analyze the likers' photos that receive like-backs. As shown in Figure 11, among all users who liked back their likers, 78.99% of them provide one like-back, and 10.03% of them give two. By studying the position of the likers' photos that get like-backs, we discover that more than 79% of like-backs starts from the first photo (see Figure 12), and nearly 9% of liking back starts from the second one, while only 4.13% of liking back start from the tenth photo. We further analyze the content of each photo being liked back, and find that more than half of them include no hashtags, the ones with hashtags are highly random. These indicate that a #like4like publisher's liking back behavior is rather random. Namely, he does not bother to check all the photos of a liker, but simply likes the first one he sees to finish what he has promised with #like4like.

In summary, we discover that most of the #like4like publishers do *not* keep their promises; the speed of like-back of publishers is very fast; the likers who interacted with #like4like publishers before are more likely to receive like-backs; the publishers' liking back behavior is rather random meaning that they only like back to keep their promises, rather than the actual photos.

6 RELATED WORK

With OSNs becoming an indispensable part of people's life, academia starts to utilize users' data in social networks to reshape our understanding about the society. Previously, Facebook and Twitter data have been the primary source for this purpose, researchers have studied people's political affiliation [14], information consumption [2, 7, 17], mobility [1, 12, 18] and cultural differences [6]. During the past five years, with Instagram exhibiting a rapid growth, more and more researchers concentrate on understanding Instagram users' behaviors [3, 5, 8, 9, 13, 16].

Manikonda et al. [8] are among the first to study Instagram by looking into its user activities, social network structure and user generate contents. They have collected 5.6 million photos through Instagram's API and discovered a few interesting facts, such as Instagram users' social network structure is quite different from other popular social networks, including Twitter and Flickr, with respect to homophily and clustering coefficient, Instagram users are much more willing to share their locations when publishing photos and a user shares a photo every 6.5 days in average. The authors of [3] have also studied Instagram's network structure by considering not only users' relationships but also their interactions represented by comments and likes. Their discoveries include users' topical interests influence users' communication behaviors, most users exhibit limited vocabularies of hashtags and popular users are more likely to have a broader interest.

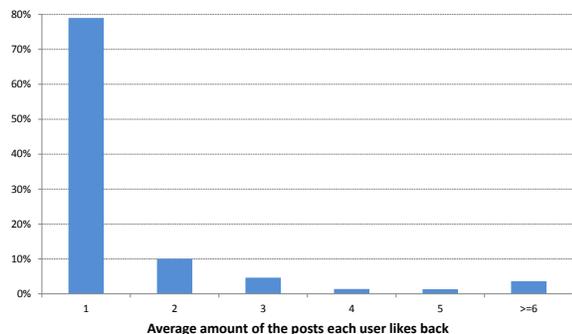


Figure 11: Average number of liked-back photos.

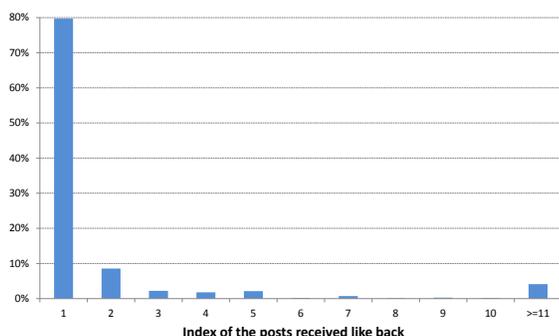


Figure 12: Distribution of the index of liked-back photos.

Besides the social network itself, the large quantity of users shared hashtags in Instagram have raised academia’s interests recently. The authors of [16] use #selfie to study the phenomenal online self-portrait convention. From a large dataset composed by more than 1 million photos, they discover that the number of #selfie has increased by 900 times from 2012 to 2014. Also, the young female users are the most prominent groups to share selfies. Moreover, they discover that selfie behaviors are correlated with countries’ cultural and socioeconomic contexts: in countries where people trust each other and feel more control over their own lives, users share fewer selfies. Mejova et al. [9] study #foodporn which is normally used to describe the appetizing pictures of users’ favorite food. As expected, the authors discovered that photos appended with #foodporn are mainly depicting high calorie food such as chocolate and cake. However, experimental results also show that #foodporn appears quite often in a healthy context, and photos of this kind also receive a higher social approval, in terms of the number of likes and comments.

7 CONCLUSION

In this paper, we have studied one of the most popular hashtags in Instagram, namely #like4like. Since #like4like implies an interaction between its publishers and likers, our analyses are centered around two research questions, i.e., whether a publisher will get more likes by appending #like4like to his photos and whether a liker will receive likes from a publisher if he likes the publisher’s #like4like photo. Through statistical analysis on our Instagram datasets, our general answer to this question is that #like4like will not provoke

likes for neither the publishers and the likers. For the first research question, we have discovered that: #like4like will not get extra likes after the year 2013; health and travel related photos are easier to get likes; users tend to like #like4like photos whose publishers are similar to them; the lifespan of a #like4like photo is very short (less than one day). For the second research question, we have shown that most of the #like4like publishers do not keep their promises by liking back their likers’ photos; for those publishers who keep their promises, they like back very fast (within 15 minutes) and the photos they like back are rather randomly chosen.

Our results in this paper can be used to understand how users behave and interact with each other in OSNs, especially in Instagram. This might lead to the creation of new hashtags to facilitate user interactions. More investigations into the datasets might also allow us build interesting social behavior models and study whether and why people behave differently in the digital and real worlds.

Moreover, hashtags are highly diverse and very popular among users in Instagram, and they can be used to indicate photos’ contents and users’ emotions, to promote events, even to reflect users’ political viewpoints. In particular, we intend to continue our research into hashtags to understand their privacy implications.

ACKNOWLEDGMENTS

Minyue Ni and Weili Han are supported by the Shanghai Innovation Action Project (No. 16DZ1100200) and NSFC (No. 61572136).

REFERENCES

- [1] Lars Backstrom, Eric Sun, and Cameron Marlow. Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity. In *WWW 2010*. ACM, 61–70.
- [2] Ismail Badache and Mohand Boughanem. Emotional Social Signals for Search Ranking. In *SIGIR 2017*. ACM.
- [3] Emilio Ferrara, Roberto Interdonato, and Andrea Tagarelli. Online Popularity and Topical Interests through the Lens of Instagram. In *HT 2014*. ACM, 24–34.
- [4] Stamatiou Giannoulakis and Nicolas Tsapatoulis. 2016. Evaluating the Descriptive Power of Instagram Hashtags. *Journal of Innovation in Digital Ecosystems* 3, 2 (2016), 114–129.
- [5] Yu-I Ha, Sejeong Kwon, Meeyoung Cha, and Jungsoek Joo. Fashion Conversation Data on Instagram. In *ICWSM 2017*. The AAAI Press, 418–427.
- [6] Lichan Hong, Gregorio Convertino, and Ed H. Chi. Language Matters in Twitter: A Large Scale Study. In *ICWSM 2011*. The AAAI Press, 518–521.
- [7] Liangjie Hong, Ovidiu Dan, and Brian D Davison. Predicting Popular Messages in Twitter. In *WWW 2011*. ACM, 57–58.
- [8] Lydia Manikonda, Yuheng Hu, and Subbarao Kambhampati. 2014. Analyzing User Activities, Demographics, Social Network Structure and User-generated Content on Instagram. *CoRR* (2014).
- [9] Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. Fetishizing Food in Digital Age: #foodporn Around the World. In *ICWSM 2016*. AAAI, 250–258.
- [10] Minyue Ni, Yang Zhang, Weili Han, and Jun Pang. An Empirical Study on User Access Control in Online Social Networks. In *SACMAT 2016*. ACM, 13–23.
- [11] Jun Pang and Yang Zhang. DeepCity: A Feature Learning Framework for Mining Location Check-Ins. In *ICWSM 2017*. The AAAI Press, 652–655.
- [12] Jun Pang and Yang Zhang. Location prediction: Communities Speak Louder Than Friends. In *COSN 2015*. ACM, 161–171.
- [13] Jun Pang and Yang Zhang. Quantifying Location Sociality. In *HT 2017*. ACM.
- [14] Marco Pennacchiotti and Ana-Maria Popescu. A Machine Learning Approach to Twitter User Classification. In *ICWSM 2011*. The AAAI Press, 281–288.
- [15] Miriam Redi, Daniele Quercia, Lindsay Graham, and Samuel Gosling. Like Partying? Your Face Says It All. Predicting the Ambiance of Places with Profile Pictures. In *ICWSM 2015*. The AAAI Press, 347–356.
- [16] Flavio Souza, Diego de Las Casas, Vinicius Flores, SunBum Youn, Meeyoung Cha, Daniele Quercia, and Virgilio Almeida. Dawn of the Selfie Era: The Whos, Wheres, and Hows of Selfies on Instagram. In *COSN 2015*. ACM, 221–231.
- [17] Jiang Yang and Scott Counts. Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. In *ICWSM 2010*. The AAAI Press, 355–358.
- [18] Yang Zhang and Jun Pang. Distance and Friendship: A Distance-based Model for Link Prediction in Social Networks. In *APWeb 2015*. Springer, 55–66.