

# What We Instagram: A First Analysis of Instagram Photo Content and User Types

Yuheng Hu, Lydia Manikonda, Subbarao Kambhampati @ Arizona State University

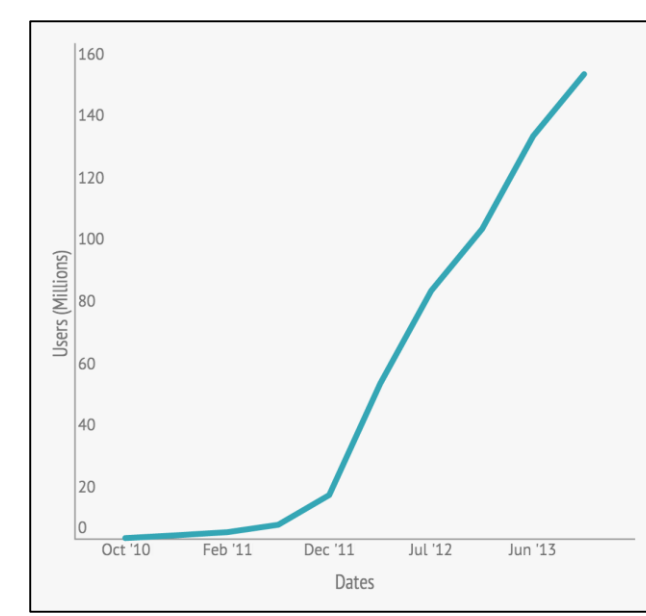
Contact: [yuheng@asu.edu](mailto:yuheng@asu.edu), [hiheng@Twitter](https://twitter.com/hiheng)

## Motivation

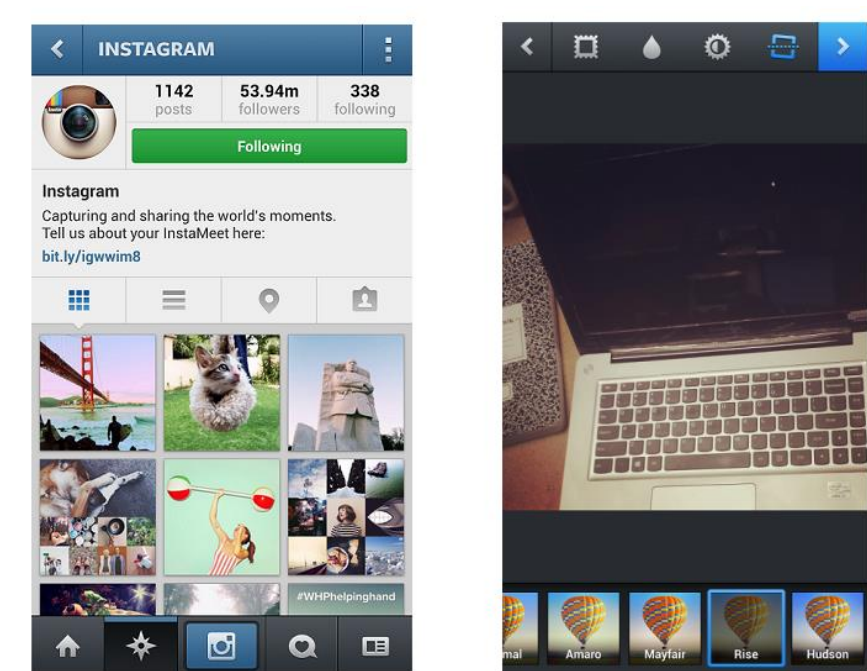


Instagram, a mobile photo (and video) capturing and sharing service, has quickly emerged as a new medium in the recent years.

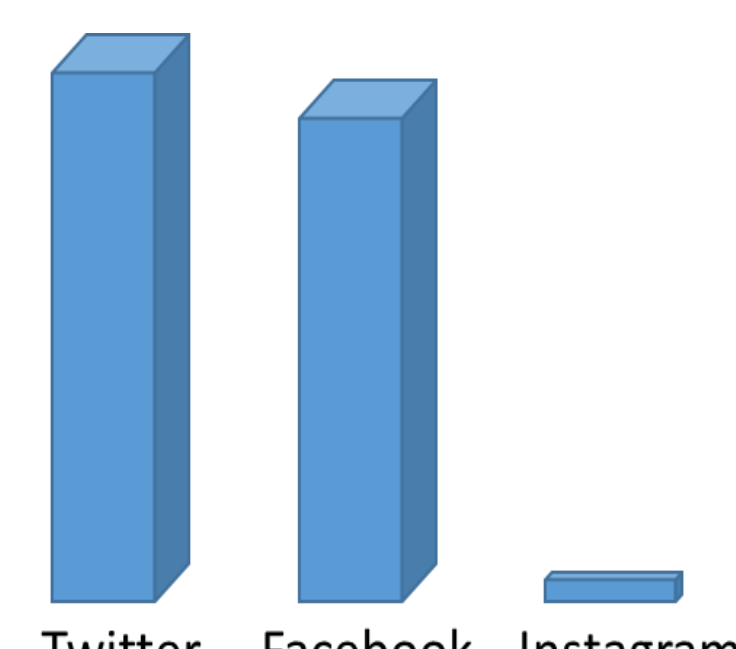
- It allows users to post photos/videos using their smartphones, apply different manipulation tools (filters) to transform the appearance of a photo, and share them instantly on multiple platforms (e.g., Instagram, Twitter, Facebook)
- It allows users to add captions, hashtags, geolocations to describe the photos and videos, and tag or mention other users before posting them
- It also provides similar social connectivity as Twitter that allows a user to follow any number of other users.



Growth of Instagram (from nitrogr.am 2014)



Instagram Interface



Num. of research papers

Despite its popularity, little research has been focused on Instagram.

- We advocate that Instagram deserves attention from the research community that is comparable to the attention given to Twitter and other social media platforms

Research question in this paper:

- What types of photos and videos do people usually post on Instagram?
- What are the differences between users in terms of their posted photos?
- How are these differences between users' photos related to other user characteristics, such as the number of followers?

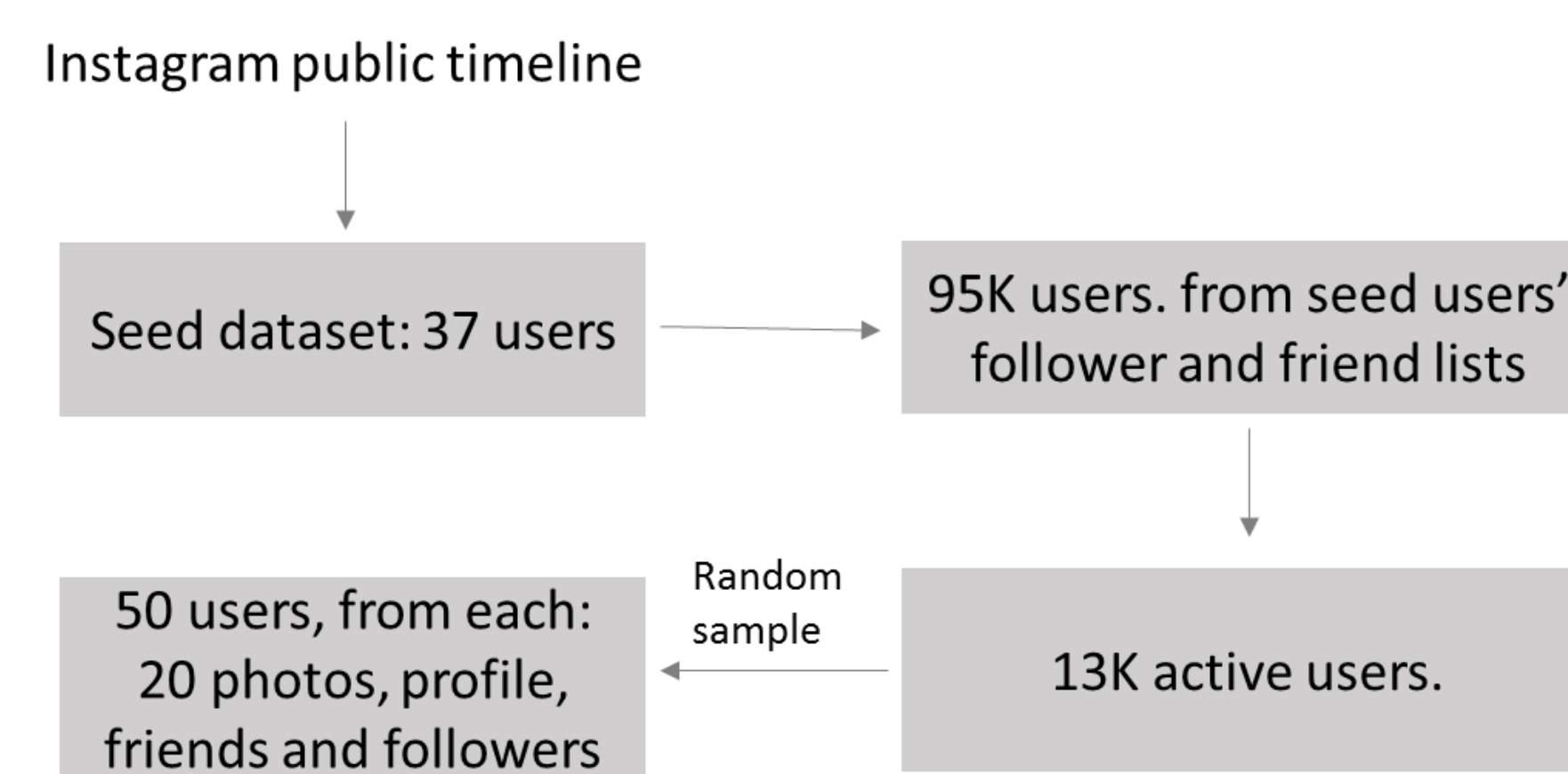
Value of this study:

- Help us gain deep insights about social, cultural and environmental issues about people's activities (through the lens of their photos)

**A picture is worth a thousand words!!**

## Dataset

- Goal: To get a random sample of Instagram users and their public photos



No brands, orgs or spammers; has at least 30 friends, 30 followers, and 60 photos

## Methods

We address research questions using manual coding, computer vision, and machine learning

### Coding scheme (based on 200 photo sample)

Computer vision techniques to get an overview of what categories exist on Instagram in an efficient manner

- Extract Scale Invariant Feature Transform (SIFT) features
- Obtain codebook vectors using *Image Vector Quantization approach*
- Apply *k*-means to obtain 15 *initial* categories of photos

Human coders to examine initial coding scheme

- Adjust 15 categories independently
- Exchange and resolve conflicts in categorizations during a discussion session, finally lead to 8 categories

Category	Exemplary Photos
Friends (users posing with others friends; At least two human faces are in the photo)	
Food (food, recipes, cakes, drinks, etc.)	
Gadget (electronic goods, tools, motorbikes, cars, etc.)	
Captioned Photo (pictures with embed text, memes, and so on)	
Pet (animals like cats and dogs which are the main objects in the picture)	
Activity (both outdoor & indoor activities, places where activities happen, e.g., concert, landmarks)	
Selfie (self-portraits; only one human face is present in the photo)	
Fashion (shoes, costumes, makeup, personal belongings, etc.)	

### Coding process

- Two coders categorize the rest 800 photos based on their main themes and their descriptions and hashtags independently (kappa=0.75)
- Single category to each photo
- Third-party judges to view the unresolved photos

### User type clustering

- 8-dimensional vector for each user. Each dimension represents the proportion of user's photos in the corresponding category
- Apply *k*-means clustering

## Results

- Q1: What kind of photos do people usually post on Instagram?

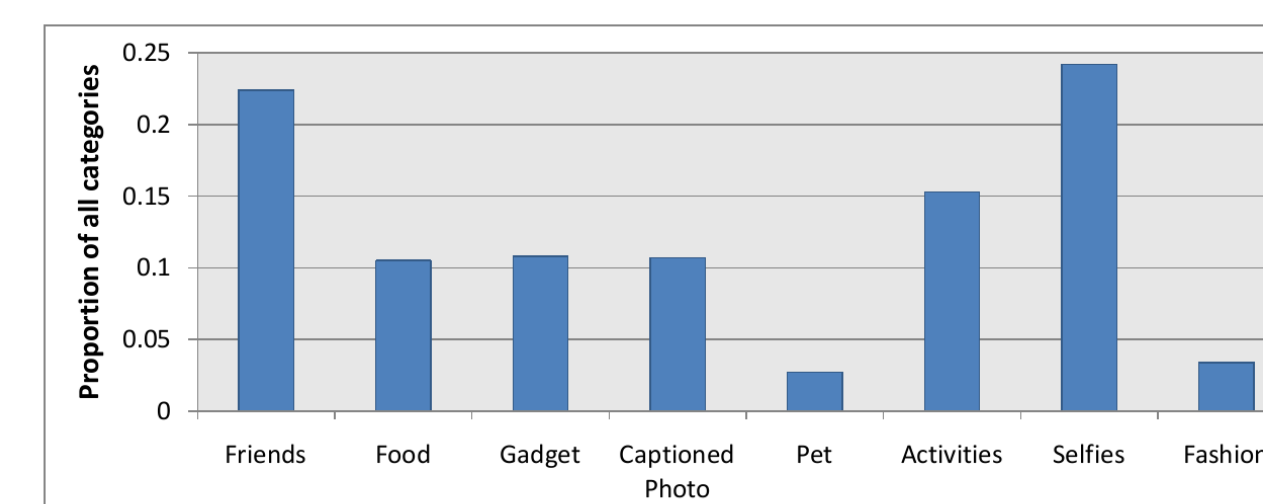


Figure 2: Proportion of Categories

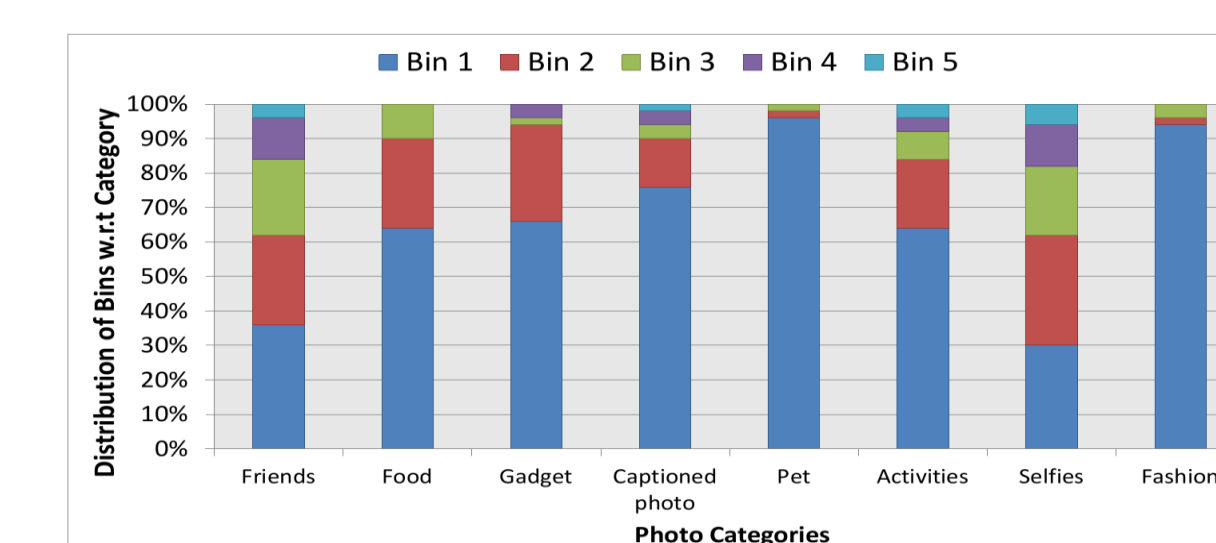


Figure 3: Proportion of users w.r.t content categories. Bin1 contains 0-2 photos; Bin2 contains 3-5 photos; Bin3 contains 6-8 photos; Bin4 contains 9-11 photos; Bin5 contains 11 photos.

Instagram is mostly used for self promoting and social networking with their friends.

Pet and Fashion has high  $SD=0.5$  compared to Selfies  $SD=0.11$  and Friends  $SD=0.124$

## Results

- Q2: How do the users differ based on the type of images they post?

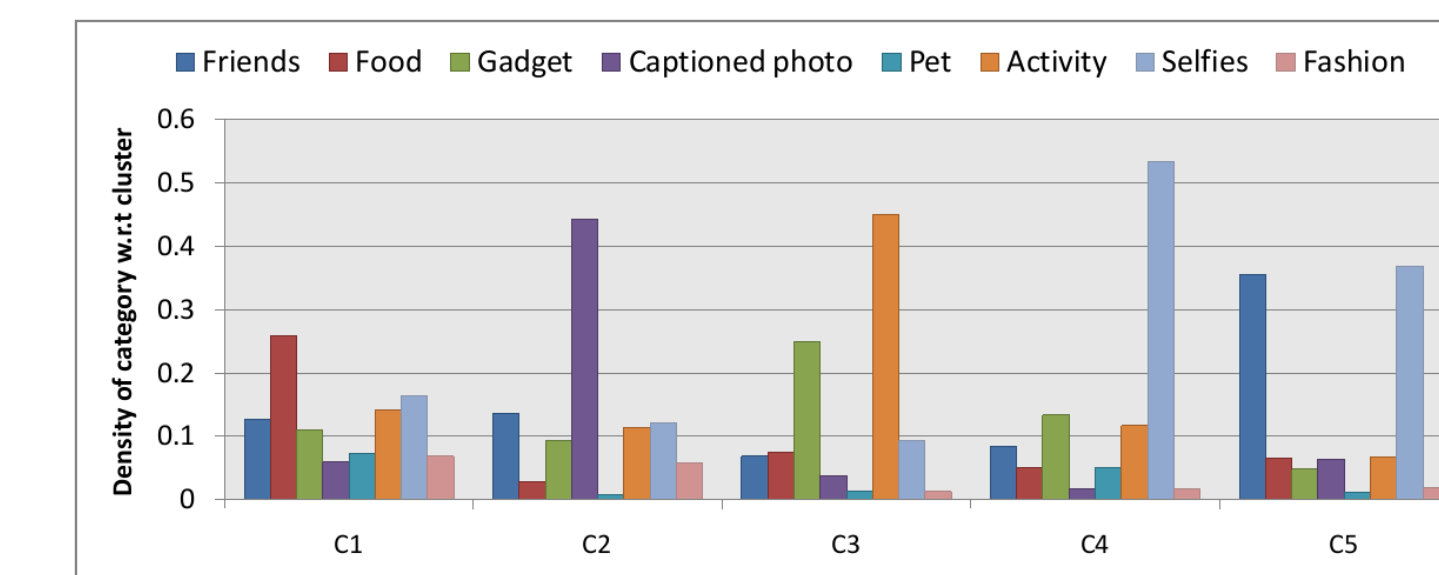


Figure 4: Clustering users based on the categories of their photos. C1 to C5 represent five different user clusters. C1 (n=11, 22%), C2 (n=7, 14%), C3 (n=7, 14%), C4 (n=3, 6%), and C5 (n=22, 44%)

People in C5 (22 users in total) care about their friends as seriously as caring about themselves, by posting nearly equal number of photos from both categories (while ignoring the other categories)

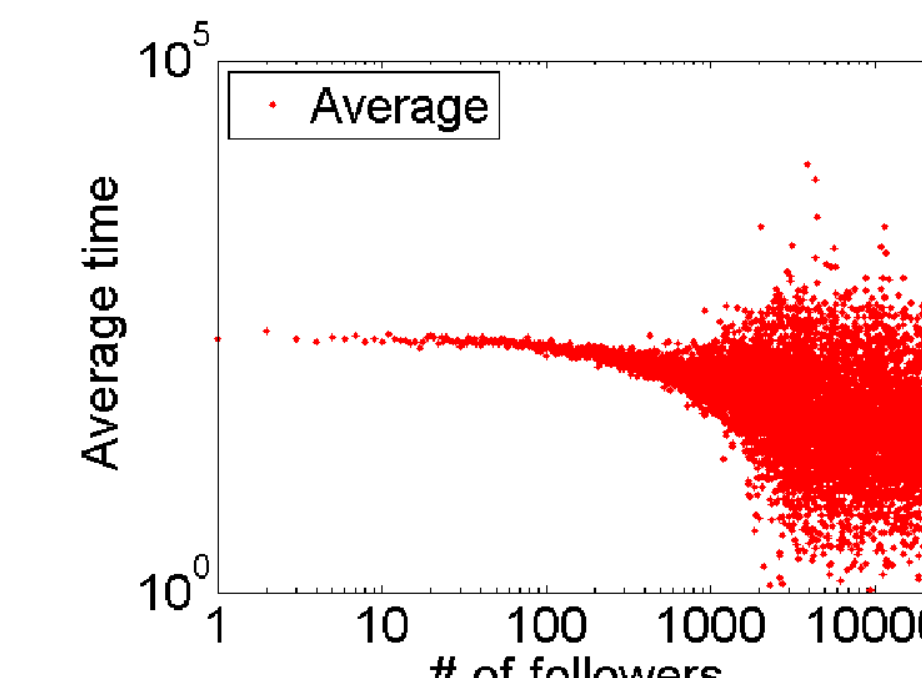
- Q3: How are these differences between users' photo content related to user's number of followers?

- Do "selfies-lovers" (C4) attract significantly more followers than common users in C1?

We perform *t*-test on the *follower distributions* from different user clusters. We find that all the other types of users agree with the null hypothesis that followers are independent of the user clusters (two-tailed *t*-test;  $p$ -value = 0.171). So #followers is **NOT** correlated with types of users

## Other analysis results

Based on 2<sup>nd</sup> Instagram dataset: 370K users, 5M photos, 2M comments



Average time between two posts by a typical user on Instagram is 6.5 days



Follow, love, life are the top 3 unigrams in user's bio

Only 14.9% of reciprocal relationship between Instagram users, given that users explicitly ask others to follow them (such request is mentioned in their bio)

## Conclusions

- To our knowledge, this is the first paper that conducts such analysis on Instagram data
- Our analysis shows that there are largely 8 different types of photo categories on Instagram. Based on the content posted by users, this analysis derives 5 different types of users (or user clusters)
- We also showed that there is no direct relationship between the number of followers and the type of users, through statistical significance tests

Instagram	Twitter (Kwak et al.)	Flickr (Cha et al.)
14.9%	22.1%	68%