Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Tweeting the Mind and Instagramming the Heart: Exploring Differentiated Content Sharing on Social Media

Lydia Manikonda Venkata Vamsikrishna Meduri Subbarao Kambhampati
Department of Computer Science, Arizona State University, Tempe AZ 85281
{lmanikon, vmeduri, rao}@asu.edu

Abstract

Understanding the usage of multiple Online Social Networks (OSNs) is of significant research interest as it helps in identifying the distinguishing traits of each social media platform that contribute to its continued existence. A comparison between two OSNs is particularly useful when it is done on the representative set of users holding active accounts on both the platforms. In this research, we collected a set of users holding accounts on both Twitter and Instagram. An extensive textual and visual analysis on the media content posted by these users reveals that these platforms are indeed perceived differently at a fundamental level with Instagram engaging more of the users' *heart* and Twitter capturing more of their *mind*. These differences get reflected in the linguistic, topical and visual aspects of the user posts.

1 Introduction

Twitter and Instagram are popular microblogging services with many users having active accounts on both these sites (or platforms) (Lim et al. 2015; Chen et al. 2014). While research has recognized immense practical value in understanding the user behavioral characteristics on these platforms separately, there is no existing research that has examined *how the content posted by the same set of individuals differs across these two platforms*. Instagram is a photo-sharing application whereas Twitter emerged as a text-based application which currently lets users post both text and multimedia data. Of particular interest is the question of *why and how a particular individual uses these two sites when both of them are similar in their current functionalities.* We aim to answer this question by analyzing content from the same set of individuals across these two popular platforms and quantifying their posting patterns. We focus on ordinary users who are neither celebrities nor popular organizations. By leveraging Natural Language Processing and Computer Vision techniques, we present some of the first qualitative insights about the types of trending topics, and social engagement of the user posts across these two platforms. Analysis on the visual and linguistic cues indicates the dominance of personal and social aspects on Instagram and news, opinions and work-related aspects on Twitter. As a part of the linguistic analysis, we observed that Instagram is largely a sphere of positive personal and social information where as Twitter is primarily a news sharing media with higher negative emotions shared by users. We see the same set of users posting remarkably different categories of visual content – predominantly eight categories on Instagram and four categories of images on Twitter. An extended analysis on textual and visual content is presented in the archive version (Manikonda, Meduri, and Kambhampati 2016) of this paper.

Background: Twitter has been explored extensively with respect to the content (Honey and Herring 2009), and language (Hong, Convertino, and Chi 2011). It is established that Twitter is primarily a news medium (Kwak et al. 2010). Research on Instagram has focused mostly on understanding the user behavior through analyzing color palettes (Hochman and Schwartz 2012), categories (Hu, Manikonda, and Kambhampati 2014), filters (Bakhshi et al. 2015), etc. On the other hand, it has been of significant interest to the researchers to investigate the behavior of a user (Benevenuto et al. 2009), mapping same user accounts (Zafarani and Liu 2013), study how users reveal their personal information (Chen et al. 2012), etc all across multiple OSNs. We extend the current state of the art by examining the nature of a given user's behavior manifested across Twitter and Instagram. Close to our research is the work of Raphael et al. (Ottoni et al. 2014) that compared Pinterest and Twitter, Bang et al. (Lim et al. 2015) where six OSNs were studied to analyze the temporal and topical signature (only w.r.t user's profession) of user's sharing behavior but they did not focus on studying the comparative linguistic aspects and visual cues across the platforms. Here we employ both textual and visual techniques to conduct a deeper comparative analysis of content on both Twitter and Instagram.

2 Dataset

To get a set of users maintaining accounts on both OSNs, we use a personal web hosting service called *About.me* (http://about.me/). This service enables individuals to create an online identity by letting them provide a brief biography, connections to other individuals and their personal websites. Using its API, we initially crawled a set of 10,000 users and pruned users who do not maintain accounts on both the platforms. The final crawl includes 963 users with a total of 1,035,840 posts from Twitter (using the Twitter API https://dev.twitter.com/overview/api) and 327,507

Twitter Topic Vocabulary		Instagram Topic Vocabulary	
ID	Terms	ID	Terms
0	stories, international, food, web, não, angelo, já	0	#food, delicious, coffee, sunset, beautiful, happy, #wedding
1	time, people, love, work, world, social, life	1	#streetart, #brighton-graffiti, #belize, #sussex, #hipstamatic, #urbanart, #lawton
2	happy, love, home, birthday, weekend, beautiful, park	2	#fashion, #hair, #makeup, #health, #workout, #vegan, #fit
3	más, día, vía, gracias, mi, si, las	3	#instagood, #photooftheday, #menswear, #style, #travel, #beach, #summer
4	#football, #sports, #news, #art, facebook, google, iphone	4	birthday, beautiful, love, christmas, friends, fun, home

Table 1: Words corresponding to the 5 latent topics from Twitter and Instagram

posts from Instagram (using the Instagram API https://www.instagram.com/developer/). Each post in this dataset is public and the data include user profiles along with their followers and friends list, tweets (insta posts), meta data for tweets that include favorites (likes), retweets (Instagram has no explicit reshares; so we use comments as a measure of the attention the post receives), geo-location tagged, date posted, media content attached and hashtags.

3 Text Analysis

3.1 Latent Topic Analysis

To analyze the types of content posted by a user across Twitter and Instagram, we first mine the latent topics from the corpus of Twitter (aggregated posts on Twitter of all users) and corpus of Instagram (aggregated posts on Instagram of all users where we use captions associated with posts for this analysis). Topic analysis is meaningful as it is pertinent to understand the reasons behind users joining the two platforms and making posts actively. We use TwitterLDA (https://github.com/minghui/Twitter-LDA) developed for topic modeling of short text corpora to mine the latent topics.

The topic vocabulary listed for both the platforms in Table 1 indicates the unique topics for each site as well as the overlapping topics. For instance topics 0 and 4 on Instagram are similar to the topics 1 and 2 on Twitter. However, a significant difference is that Instagram is predominantly used to post about art, food, fitness, fashion, travel, friends and family but Twitter hosts a significantly higher percentage of posts on sports, news and business. Another notable difference is that the vocabulary from non-English language posts like French and Spanish is higher on Twitter as compared to the captions on Instagram mostly using English as the language medium. The topic distributions obtained from the two corpora listed in Figure 1 show that friends and food are the most frequently posted topics on Instagram as against sports and news followed by work and social life being popular on Twitter.

To further validate the observations made about the distinctive topical content across the two platforms, we compared the topic distributions for each individual on the two

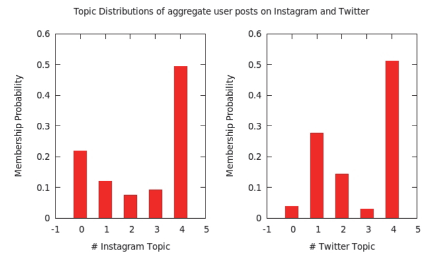


Figure 1: Topic distributions of all the user posts on Twitter and Instagram

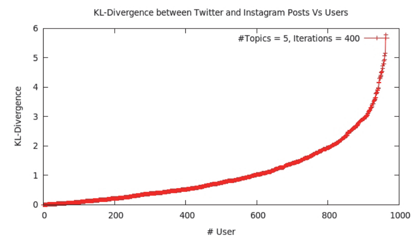


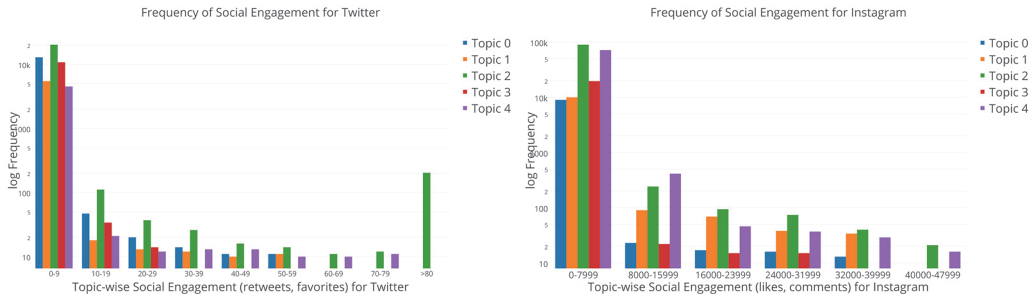
Figure 2: Sorted entropies between the topic distributions of the user posts on Twitter and Instagram

platforms by estimating the *KL-Divergence* between the topic distributions on each platform. We do this by first building a unified topic model on the combined corpus of tweets and captions of Instagram posts. The unified topics are listed in the description of Figure 3. The resultant entropy plot in Figure 2 shows a significant fraction of the users posting distinct content on the two platforms. This distinction is statistically significant with an estimated p-value $< 10^{-15}$ for each user.

3.2 Social Engagement

Since our findings revealed that the topics across the two platforms are significantly different, we wanted to investigate how the posts made by the same user engage other individuals on the two sites. We define the social engagement as the attention received by a user's post on the social media platform. It can be quantified in various ways varying from the sum of likes and comments on Instagram to the sum of favorites and reshares on Twitter. For each topic in the unified topic model for both Twitter and Instagram, the logarithmic frequency of posts is plotted against the magnitude of social engagement that is binned to discrete ranges. The results are shown in Figure 3.

An interesting observation is that the socially engaging topics in the combined model are same as the overlapping topics from the topic models built in isolation on the Twitter and Instagram posts (Figure 1). The dominating topic on Twitter is about sports, news and business and on Instagram it is about love, fashion and food. Surprisingly, we found that the overlapping topics (Topics 2 and 3) focusing on social and personal life fetched predominant social engagement on both the platforms. Though the dominating topic (commonly



(a) *Twitter* (b) *Instagram*

Figure 3: Social Engagement Vs Post Frequency where the topics are – Topic 0:{people, life, world, social, app, game, business}, Topic 1:{stories, artists, #lastfm, level, #football, #sports, news}, Topic 2:{birthday, beautiful, work, weekend, park, dinner, christmas}, Topic 3:{ yang, run, #fitness, #runkeeper, #art, sale, #menswear}, Topic 4:{#instagood, #photooftheday, #love, más, #fashion, #travel, #food}

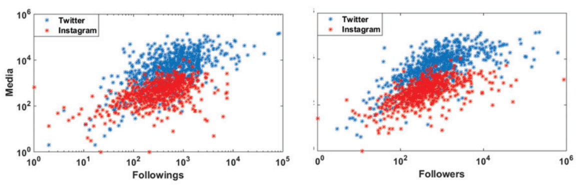


Figure 4: Distributions of Followers/Followings vs Media

posted) and the most engaging topic (commonly liked) happen to be the same on Instagram, this phenomenon is not applicable to Twitter.

A notable difference between the platforms with respect to social engagement is that the magnitude of attention received for Instagram posts is significantly higher than the level of attention received on Twitter. We can see this from the ranges plotted on the x-axes in Figure 3. This observation is consistent regardless of the activity level of a user. Even when a user is more active (Figure 4) on Twitter than Instagram, the observation of higher social engagement on Instagram on an absolute scale holds. A possible explanation to this is that the users on Twitter use it as a news source to read informative tweets but not necessarily all of the content that is read will be "liked".

On average, there are 30% more hashtags for a Twitter post compared to an Instagram post (Pearson correlation coefficient = 0.34 between distributions with p-value $< 10^{-15}$). This may also indicate that on Instagram since the main content is image, textual caption may not receive as much attention from the user.

3.3 Linguistic Nature

To characterize and compare the type of language used on both platforms, we use the psycholinguistic lexicon LIWC ((http://liwc.wpengine.com/)) on the text associated with Twitter posts and Instagram posts. We obtain measures of attributes related to user behavior – *emotionality* (how people are reacting to different events), *social relationships* (friends, family, other humans) and *individual differences* (attributes like bio, gender, age, etc).

	Platform	
	Twitter	Instagram
Emotionality		
Negemo	0.60	0.49
Posemo	0.19	0.19
Social Relationships		
home	0.15	0.30
family	0.14	0.21
friend	0.05	0.1
humans	0.17	0.21
Individual Differences		
work	0.81	0.5
bio	0.6	0.93
swear	0.08	0.06
death	0.07	0.04
gender	0.16	0.2

Table 2: Linguistic attributes across Twitter vs Instagram. Each value indicates the fraction of a post belonging to the corresponding attribute

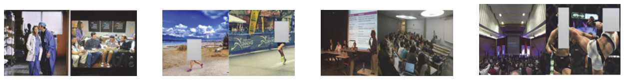
It is clear from Table 2 that posts on Twitter have more negative emotions and contain more work-related and swear words. In contrast, positive social patterns are more evident on Instagram. By relating these results to the topic analysis results in the previous section, we note that on Instagram users share more light-hearted happy personal updates.

4 Visual Analysis

This section develops a better understanding of the types of photos individuals post on Twitter in comparison with their Instagram posts. To achieve this we employ computer vision techniques mainly in terms of *visual categories* (kinds of photos).

4.1 Visual Categories

We further investigated if the visual categories of the posts made on Twitter and Instagram are different. We first sampled two sets of 5000 images from both platforms separately. Using the OpenCV library (http://opencv.org/) on these two datasets, we extracted Speeded Up Robust Features (SURF) for each image. We used the vector quantization approach



(a) (b) (c) (d)

Figure 5: Subcategories of *activity*: a) TV shows, b) Running, c) Conferences, d) Live shows

(a) (b) (c)

Figure 6: Subcategories of *captioned photos*: a) Snapshots, b) Memes, c) Quotes

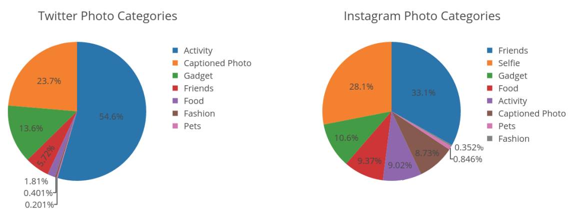


Figure 7: Photo categories on Twitter vs Instagram

on these features that eventually converted each image into a codebook format. Using the codebook, we clustered images using k-means algorithm (best value of k is found by SSE (Sum of Squared Error)) which are refined and considered as the overall visual themes or categories.

Visual categories on Instagram agree with our previous work (Hu, Manikonda, and Kambhampati 2014) which detected eight different categories of images. We tried to categorize the Twitter images into the same format as Instagram images. This shows that there are four prominent cluster categories on Twitter. Figure 7 shows that the percentage of photos in the activity category outnumbered any other category followed by captioned photos. To better understand the kinds of activities and captions shown in these two sections, we sampled around 200 images and asked the two researchers to label them manually into different sub-categories. Figure 5 indicates the most popular subcategories in the *activity* category – news, events (football games, concerts, conferences) and races. Figure 6 indicates that majority of the captioned photos are snapshots, memes, and quotes or opinions. These categories show that the topics of photos on Twitter are mainly related to news, opinions or other general user interests. In contrast, on Instagram users seem to mainly share their joyful and happy moments of their personal lives.

5 Conclusions

In this paper, we presented a detailed comparison of the textual and visual analysis of the content posted by the same set of users on both Twitter and Instagram. Some of the insights obtained from linguistic analysis reveal the fundamental differences in the thinking style and emotionality of the users on these two platforms and how the posts receive varying degrees of attention as per the underlying topics. The visual analyses with respect to categories and color palettes indicate that the pictures posted on Instagram contains more selfies and photos with friends where as Twitter contains more about user opinions in the form of captioned photos – memes, quotes, etc. We observed that the differences are deeply rooted in the very intention with which users post on these platforms with Twitter being a venue for serious posts about news, opinions and business life while Instagram serves as the host for light-hearted personal moments and posts on leisure activities. Interestingly, user posts on Instagram seem to receive significantly more attention than Twitter.

Acknowledgements: This research is supported in part by a Google research award, the ONR grants N00014-13-1-0176, N00014-13-1-0519 and N00014-15-1-2027, and the ARO grant W911NF-13- 1-0023. We thank Reza Zafarani for providing us the initial seed set of users.

References

Bakhshi, S.; Shamma, D.; Kennedy, L.; and Gilbert, E. 2015. Why we filter our photos and how it impacts engagement. In *ICWSM*.

Benevenuto, F.; Rodrigues, T.; Cha, M.; and Almeida, V. 2009. Characterizing user behavior in online social networks. In *IMC*.

Chen, T.; Kaafar, M. A.; Friedman, A.; and Boreli, R. 2012. Is more always merrier?: A deep dive into online social footprints. In *WOSN*.

Chen, Y.; Zhuang, C.; Cao, Q.; and Hui, P. 2014. Understanding cross-site linking in online social networks. In *SNAKDD*.

Hochman, N., and Schwartz, R. 2012. Visualizing instagram: Tracing cultural visual rhythms. In *ICWSM*.

Honey, C., and Herring, S. 2009. Beyond microblogging: Conversation and collaboration via twitter. In *HICSS*, 1–10.

Hong, L.; Convertino, G.; and Chi, E. 2011. Language matters in twitter: A large scale study. In *ICWSM*.

Hu, Y.; Manikonda, L.; and Kambhampati, S. 2014. What we instagram: A first analysis of instagram photo content and user types. In *ICWSM*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *WWW*.

Lim, B. H.; Lu, D.; Chen, T.; and Kan, M.-Y. 2015. #mytweet via instagram: Exploring user behaviour across multiple social networks. In *ASONAM*.

Manikonda, L.; Meduri, V.; and Kambhampati, S. 2016. Tweeting the mind and instagramming the heart: Exploring differentiated content sharing on social media. *CoRR* abs/1603.02718.

Ottoni, R.; Casas, D. L.; Pesce, J. P.; Meira Jr., W.; Wilson, C.; Mislove, A.; and Almeida, V. 2014. Of pins and tweets: Investigating how users behave across image- and text-based social networks. In *ICWSM*.

Zafarani, R., and Liu, H. 2013. Connecting users across social media sites: A behavioral-modeling approach. In *KDD*.