Characterising and detecting sponsored influencer posts on Instagram

Koosha Zarei, Damilola Ibosiola, Reza Farahbakhsh, Zafar Gilani, Kiran
Garimella, Noel Crespi, Gareth Tyson

▶ **To cite this version:**

Characterising and Detecting Sponsored Influencer Posts on Instagram

Koosha Zarei*, Damilola Ibosiola†, Reza Farahbakhsh*, Zafar Gilani†, Kiran Garimella‡, Noël Crespi*, Gareth Tyson†

*Institut Polytechnique de Paris, Télécom SudParis France. {koosha.zarei, reza.farahbakhsh, noel.crespi}@telecom-sudparis.eu
†Queen Mary University of London, United Kingdom. {d.i.ibosiola, z.gilani, gareth.tyson}@qmul.ac.uk
‡Massachusetts Institute of Technology. garimell@mit.edu

Abstract—Recent years have seen a new form of advertisement campaigns emerge: those involving so-called social media *influencers*. These influencers accept money in return for promoting products via their social media feeds. Although this constitutes a new and interesting form of marketing, it also raises many questions, particularly related to transparency and regulation. For example, it can sometimes be unclear which accounts are officially influencers, or what even constitutes an influencer/advert. This is important in order to establish the integrity of influencers and to ensure compliance with advertisement regulation. We gather a large-scale Instagram dataset covering thousands of accounts advertising products, and create a categorisation based on the number of users they reach. We then provide a detailed analysis of the types of products being advertised by these accounts, their potential reach, and the engagement they receive from their followers. Based on our findings, we train machine learning models to distinguish sponsored content from non-sponsored, and identify cases where people are generating sponsored posts without officially labelling them. Our findings provide a first step towards understanding the under-studied space of online influencers that could be useful for researchers, marketers and policymakers.

Index Terms—Influencers, Instagram, Social Media, User Behaviour Analysis, Sponsored Content

I. INTRODUCTION

Social media has become an active part of billions of people's lives. Initially intended as a method to interact amongst friends, it has since become a platform used heavily for marketing. This is dominated by players such as Instagram who allow third parties to micro-target adverts at users matching certain criteria. This newfound ability has turned social media into a multi-billion dollar industry. However, in recent years a new breed of social advertisement has emerged. Driven by the rise of social media celebrities, so-called *influencers*, we have witnessed various companies marketing directly through celebrity endorsements (rather than via official advertisement channels on the platform). For example, this may involve a celebrity posting an image of them wearing a particular item of clothing or driving a particular car. Whereas endorsement by celebrities is nothing new, social platforms have enabled this to take place on an industrial scale. The proliferation of these activities, via the use of Influencer Markets,[1] has also dramatically increased the scale and accessibility of these types of promotion.

A particularly interesting trend is the rise of *nano* influencers [1]. These users may only have a small number of followers, but often have reach into highly targeted audiences. This has radically increased the number of accounts actively promoting products, thereby making the industry challenging to regulate. This is because such social media posts are highly opaque, often making it impossible to identify the difference between paid promotions, and genuine personal endorsements. This distinction is important, as personal endorsements do not fall under the same stringent regulatory framework as paid sponsorship. To address this, standards bodies such as the UK's Advertising Standards Authority (ASA) have created a set of rules asking that such post are explicitly tagged with relevant hashtags (*e.g.* #ad) [2]. Similarly platforms such as Instagram, have allowed such information to be formally embedded within posts via their user interface. To date, however, the research community has limited knowledge of how these practices are employed, or even how influencer marketing takes place more generally.

This paper strives to fill this gap, and makes two contributions. *First*, we present a broad characterisation of influencers on Instagram, and formally quantify their behavioural attributes. Although inherently of value, we also argue that this has a number of implications for emerging regulation in this sector. To this end, we collect a large-scale Instagram dataset consisting of 12k influencers (§III). We gather both Instagram posts and stories[2] from users who attach advert-related hashtags to their material.

With this dataset, we proceed to explore influencer characteristics (§IV). We see a range of promotion activity taking place, ranging from "mega" influencers with over 1 million followers to "nano" influencers with as little as 500 followers. We find a range of products being produced with health & beauty taking the top spot, followed by services, clothing and food featuring prominently. Naturally, top mega influencers garner the greatest attention, as measured via likes and comments. Yet we find that small "nano" influencers sustain attention more effectively. For example, we find that a far larger fraction of nano influencers' followers engage with posts, thus forming a denser social network (whereas, in absolute terms, higher popularity celebrities gain more likes).

---

[1]https://influencermarketinghub.com/influencer-marketing-2019-benchmark-report

[2]Stories are time limited posts that automatically delete after 24 hours.

This suggests that these different types of influences offer marketers very different styles of promotion. We also observe cases where users are creating sponsored posts, yet failing to attach appropriate labels to inform audiences that they are receiving payment. This not only has the potential to harm vulnerable users (*e.g.* children) but also raises legal questions.

This drives us to make our *second* contribution (§V), where we strive to build automated tools to detect sponsored posts that fail to explicitly tag themselves. We propose a Contextual LSTM Neural Network classifier which considers the text content and other metadata to identify whether a post is sponsored or not. We obtain an accuracy of 89%, and we identify a set of prominent predictive features, *e.g.* use of language, the inclusion of URLs and commentary. Our findings have important implications for social media companies, marketers, researchers, and regulators wishing to better understand the behaviour of influencers.

II. BACKGROUND & RELATED WORK

The concept of social influence is a well studied topic. It has long been known that an elite of *influencers* can have a disproportionate impact on social structures [3], [4]. This has subsequently led to the rise of *influencers*, who monetise their influence via sponsorship payments. The premise is that, via electronic word of mouth, brands can be promoted through influential endorsements, akin to traditional sponsorship deals [5]. More formally, sponsored influencers are "third party individuals who significantly shape the customer's purchasing decision, but may never be accountable for it" [6].

There have been a number of recent studies looking at influencer marketing, primarily within the business research community. These have followed a number of methodologies, including interviews with marketers [7], [8] and customers [9], as well as qualitative studies of specific influencers [10]. There have been further preliminary studies of the perception of influencers that, for example, found people to view influencers as smart and ambitious [11], [12].

There have also been a small set of studies that rely on computational datasets. Lahuerta *et al.* [13] studied the role of marketing influencers in the Japanese automotive sector. They collected 30K tweets based on relevant keywords. They found, amongst other things, that the sentiment and length of influencer tweets has a positive effect on popularity. Although highly relevant to our work, it is unclear how these trends generalise across other brands and large-scale influencer networks. Closest to our work is [14], which explored the follower graphs of influencers on Instagram. They targeted 218 specific social influencers (with 8.9M followers) as registered via Popular Pays, a platform used to connecting brands to influencers. These were then compared against 948 Instagram users with fewer than 10K followers. They studied the social graph properties of these users, to find that influencers have large numbers of followers, and even tend to interconnect amongst themselves. They also found that users often follow multiple influencers. Although highly applicable, we argue that these targeted studies likely only capture a small slice of the overall influencer market. They have a particular bias towards certain types of highly popular influencers, and ignore a potentially larger population of micro influencers. To the best of our knowledge, this is the first large-scale study of influencer activities.

III. DATA COLLECTION METHODOLOGY

We start by explaining our data collection methodology. Our data collection covers both Instagram posts and the recently introduced stories feature.

A. Data Collection

Our data collection activities took place over four phases:

Phase 1: Hashtags. It is first necessary to obtain a large list of influencer accounts. One approach would be to manually curate this set, however, this would limit us to a small set of influencers, largely dominated by well known celebrities who are easy to identify. Hence, we compile the list by crawling all posts attached to a set of influencer-related hashtags. We turn to the UK's Advertising Standards Agency [2], which states that influencers should use the #ad, #advert or #sponsored hashtags in *any* posts that have been paid for. We expand this list with #advertising, #giveaway, #spon and #sponsor [15].

Phase 2: Post & Stories Collection. We then use the official Instagram API [16] to gather all posts and stories that include any of the above hashtags. Note that stories are similar to normal posts, yet they are automatically deleted after 24 hours (akin to Snapchat posts). Hashtag Engine is used [17] with a maximum of 30 unique hashtags. This API returns public photos and videos that have been tagged with specific hashtags. Our crawl for posts and stories ran between Sep 2018 and April 2019. This process identifies 12K accounts that have posted using the previously mentioned hashtag.

Phase 3: Account Collection. Although the above yields a substantial body of posts and stories, we are primarily interested in gathering data on a per-influencer basis. Hence, we next extract all accounts identified from the Phase 2 dataset and begin dedicated monitoring for all posts and stories generated by those users. (*i.e.,* influencers). This covers all posts, reactions and stories from those accounts from July 2019 to August 2019. In this step, we collect 19.7K posts, 63K stories, 3.1M comments, and 27M likes (generated by the 12K user accounts from Phase 2). Note that they contain a mix of both sponsored (16%) and non-sponsored (84%) entities. For each post, we collect the image, comments, likes and public profile information of the user, as well as any other users who reacted to the post. For each story, we collect the equivalent information, although we cannot collect likes (as these are not available in stories). Each sponsored post is also tagged with the product being advertised, and the category of advertiser. In total, we have 35K posts, 99K stories, 3.1M comments, and 27M likes generated by 12K users.

Phase 4: Categorization. Once we have collected the posts and stories, it is necessary to tag explicitly which are considered sponsored. We take a simple approach. If a post is tagged

with one of the above hashtags, we assume that it is sponsored. In the case of Instagram stories, there is explicit metadata which tells us if it is sponsored (the Paid Partnership tag). Hence, for stories we rely on this metadata item (rather than hashtags). Note that this excludes posts that are sponsored, yet the user does not add the appropriate hashtag. We revisit this issue in §V.

Ethics. In line with Instagram policies as well as user privacy, we only gather publicly available data that is obtainable from Instagram. Whereas we do analyse the content of influencer posts, we do not inspect the content of comments submitted by non-influencer users.

B. Data Validation

A natural risk is that a subset of the posts containing the curated hashtags may be generated by users who are not influencers. Although it is impossible to entirely discount this at scale, we further perform manual annotation to validate the general correctness of our data. To validate our dataset, we manually looked at the profiles of the influencers to verify if they were really promoting sponsored content.

All users with above 10K followers are checked, confirming that they were all correctly tagged as posting sponsored content. We further check 25% (2K) of all influencers with under 10K followers. We find that we the above approach yields 97.6% accuracy: just 48 accounts were incorrectly classified as influencers. Note that the above only checks if a user account has one more truly sponsored posts. To provide further confidence we randomly select 500 influencers and check *all* of their posts. Around 80% of sponsored-post are correctly classified as the sponsored content (based on the hashtags previously mentioned). We filter any incorrectly identified influencers.

IV. CHARACTERIZING INFLUENCERS

We begin by exploring influencers' follower counts and engagement levels (comments and likes), before profiling the types of products promoted.

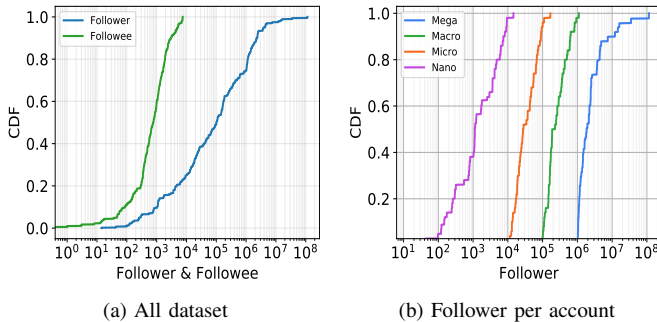(a) All dataset   (b) Follower per account

Fig. 1: Follower counts: (a) CDF of followers and followees of all the influencers in our data. (b) CDF of number of followers per account separated in groups.

A. How popular are influencers?

We take inspiration from past work [18], and begin our analysis of influence by looking at follower counts. This is a natural indicator of influence as it captures the upper-bound of people to whom posts are directly pushed. Figure 1(a) presents the cumulative distribution function (CDF) of the follower and followee counts of the influencers in our dataset.

Unsurprisingly, we see a sizeable fraction of extremely popular accounts. 35% of users have over 100K followers, and 17% posses over 1M. These conform to the common interpretation of influencers. More surprising, however, is the presence of a long tail: 37% of accounts have fewer than 10K influencers, with 15.5% even having below 1K. At first, we suspected that this may be caused by miscellaneous use of the advert-related hashtags. However, upon manual inspection, we confirm that these are indeed influencers. For instance, @lesya_9_9 with 500 followers has promoted over 10 times more regularly than @blogging_with_tiffany who has fewer than 100 followers. This reveals a growing set of small-scale influencers who promote products, despite their low follower counts. In other words, *influencers are not just celebrities*: they appear to encompass a morass of different account types. For context, we can contrast these results with the followee count (Figure 1(a)) which are, on average, far lower than follower numbers. Whereas the median follower count is 23.7K, it is just 770 for the number of followees.

TABLE I: Influencer Profile Characteristics

Influencer Category	#Followers Category	Avg. follower	Avg. followee	Avg. mediacount	% of verified
Mega	$\geq 1M$	5.8m	845	9.1k	82%
Macro	$< 1M$ & $\geq 100K$	257k	1.3k	3.1k	22%
Micro	$< 100k$ & $> 10K$	32k	1.9k	1.8k	4%
Nano	$< 10K$	1.5k	0.9k	597	0.5%

The key profile characteristics of influencers (full timeline).

Based on the above findings, we categorise influencers into 4 distinct categories based on their reach (# followers). This taxonomy underpins our subsequent analysis. We term these nano, micro, macro and mega influencers. Table I presents a summary of these groups. We note that 80% of mega influencers are verified by Instagram, but in contrast, under 5% of nano and micro accounts have a blue verified icon.[3] Figure 1(b) presents the CDF of follower counts per-account, broken into these four groups. Naturally, the distributions reflect the split with nano influencers having the fewest followers. For context, a few examples of influencers (top three in terms of followers) from the each categories is shown in Table II. As previously discussed, primarily the more "popular" influencers have verified accounts. These users tend to also have more posts on average.

B. Do followers engage with influencers?

Another way to measure "influence" is to inspect engagement levels on a users' posts, *e.g.* comments, likes and mentions. Previous studies [18] have argued that these levels can

[3]Users on Instagram can get verified badge as low as 500 followers. However, that account must represent a well-known, highly searched for person, brand or entity [19]

TABLE II: Examples of influencers

Category	Username	Section	#post	#follower	#followee	#verified	#url
Mega	@kendalljenner	Fashion	3K	116M	203	✓	-
	@vanessahudgens	Fashion	3.2K	36M	1.1K	✓	-
	@brentrivera	Lifestyle	1.8K	16M	395	✓	Youtube
Macro	@tonyamichelle26	Lifestyle	5.3K	937K	2.1K	-	Business Page
	@alice_gao	Design	4.1K	910K	500	-	Business Page
	@lilleejean	Beauty	700	950K	650	✓	Youtube
Micro	@charisseo_	Fashion	1.7K	98K	600	-	Email
	@morgbullard	Lifestyle	1.9K	97.5K	1.3K	-	Business Page
	@ginascrocca	Fashion	246	96K	1K	-	Business Page
Nano	@jaimesays	Travel	600	9K	1.9K	-	Business Page
	@aberhalloooo	Food	1.1K	9K	7K	-	Email
	@lawrence.carlyFollow	Dance	393	8.1K	1.2K	-	Business Page

be a better proxy for influencer than simply inspecting follower counts. In this section, we directly contrast engagement levels for posts that are sponsored *vs.* not sponsored.

Active attention - Comments. Figure 2(a) presents the CDF of the received comments per-post for each influencer. We separate posts into sponsored and non-sponsored posts within the influencer timelines. In almost all the cases we observe that sponsored posts receive fewer comments from the users. This suggests that these sponsored posts are of less interest than other posts. This difference is even more significant in mega influencers, where sponsored posts gained 10 times fewer comments than their non-sponsored counterparts. We also observe that the number of comments are ranked in order of Mega, Macro, Micro and Nano influencers with, unsurprisingly, Mega influencers getting over 40 times more comments than Nano.

The above analysis of absolute counts may give a misleading perspective as influencers with high follower counts (*e.g.* Mega) will obviously obtain higher comment counts. Hence, we normalize the comment count as a fraction of the follower count, and plot the results in Figure 2(b). Here, we see rather different trends with nano influencers gaining the most engagement. In other words, even though popular accounts gain more comments, less popular accounts obtain engagement from a higher fraction of their follower-base. This perhaps sheds light on why nano-influencers have recently started to gain traction, with their ability to engage more targeted populations.

We also inspect the duration before a user adds a comment on a post. Intuitively, comments that are issued shortly after a post is created might be from more engaged users. This confirms similar results to Figure 2, with Nano influencers gaining posts most rapidly than their more popular counterparts. In all the cases, non-sponsored posts gain comments earlier; the most significant difference is for Mega influencers. The median comment age for a Mega influencer's non-sponsored posts is 328 *vs.* 366.5 for their sponsored posts. That said, we observe a subtle difference between the influencer groups. In the first hour <30% of comments of all influencers are issued. After the first 10 hours, non-sponsored posts of Macro influencers receive <70% of their total comments, but sponsored posts of mega influencers get only less than 50%. The density of users who reacted to non-sponsored posts is larger, and this difference is more significant in mega influencers and less for nano influencers (plot not shown due to space constraints).

Briefly, we also note that nano influencers tend to have more consistent engagement. Whereas more popular influencers,

gain comments from many different users, nano influencers tend to get comments from the same users multiple times: 30% give more than one comment (Figure 3b).
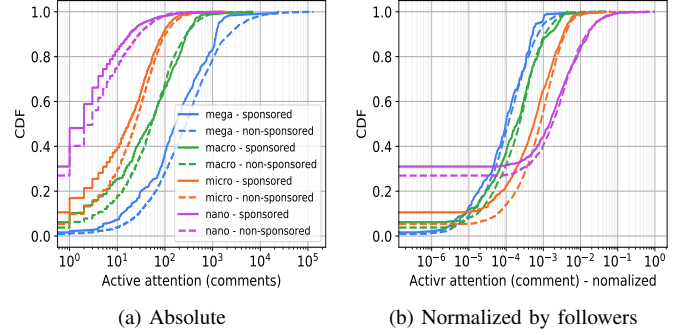

(a) Absolute   (b) Normalized by followers

Fig. 2: CDF of number of comments received per-post: (a) absolute number; (b) normalized.

Passive attention - Likes . As comments are generally the most active form of engagement, we also inspect more passive engagement via likes. Figure 3a presents the CDF of the received likes for both sponsored and non-sponsored posts among each influencer sategory. For Mega influencers, we observe 28% more likes for non-sponsored *vs.* sponsored posts. This, however, is far less for the other categories (macro, micro, and nano), with an equivalent value of 6%. In fact, we observe that sponsored posts gain marginally more likes than non-sponsored posts for nano influencers (median 56 *vs.* 47). That said, the categories exhibit broadly similar patterns to that seen in comments (with Mega gaining the most and nano gaining the fewest likes in absolute terms). Turning our attention to the normalized like count, we see that again nano influencers get more likes than the other categories. However, we do see outliers: for instance, inside nano, nearly 8% of accounts receive a large number of likes, yet nearly no comments (all posts). To investigate this, we manually inspect these subsets of nano influencers. We find that there is a prevalence of fake profiles, apparently using bots to boost their impact.
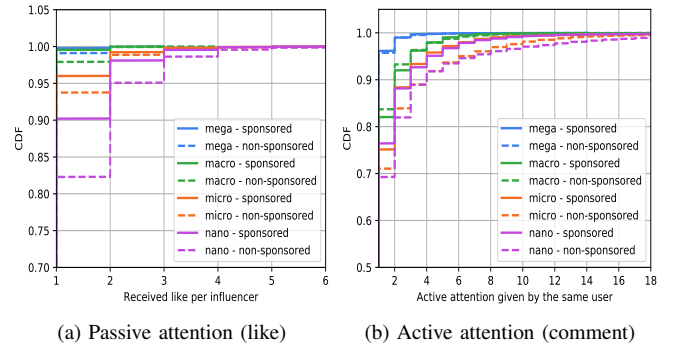

(a) Passive attention (like)   (b) Active attention (comment)

Fig. 3: (a) attracted attention CDF of number of likes received per-influencer; (b) number of comments performed by each user per-influencer

C. How often do influencers post?

We start by measuring the number of posts per influencer. Figure 4(a) presents a CDF plot of the number of sponsored *vs.* non-sponsored posts, whereas Figure 4(b) repeats the same for stories.
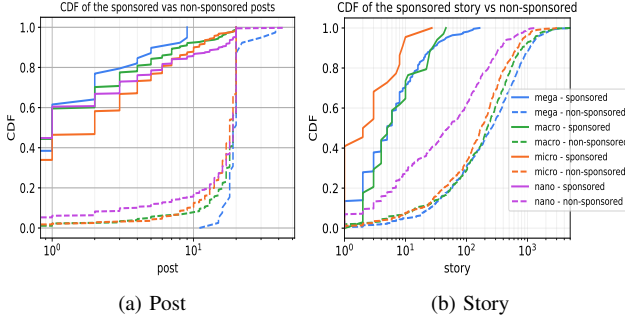

Fig. 4: CDF of the number of posts and stories published per influencer.

We observe distinct distributions, with most influencers publishing more non-sponsored posts. Only 8.3% of influencers distribute more sponsored posts compared to non-sponsored. On average, 16% of posts are sponsored with just 9.3% of influencers tagging over half of their posts as sponsored. This is anticipated as most influencer guides recommend that users keep the percentage of sponsored posts below 60%, to maintain audience engagement.

Subtle differences can also be observed between the different categories of influencer. For example, where the mega influencers on average post the most sponsored posts, they actually post the least non-sponsored posts. Of course, this might also be a product of how such influencers tag their posts. We find that no sponsored story is published by any nano influencers. This contrasts to what is observed with posts from influencers within the same nano category in Figure 4(a), where over 80% of these influencers publish ≤10 posts. This striking difference suggests that nano influencers are primarily using posts (rather than stories) to publish sponsored contents. In contrast, mega influencers tend to use stories to promote sponsored contents more regularly (compared to macro and micro influencers). For example, ≤21% of Macro and ≤3% of micro influencers publish more than 10 sponsored stories compared to over 30% for mega publishers. In general, we see that influencers across mega, macro and micro category favor the use of stories to promote sponsored content compare to post, possibly because it is cheaper to advertise via stories compared to feeds [20]. Another reason for using stories is the exclusivity *i.e.,* followers must stay engaged and must hurry to see the offer or discount code *etc.* while the story lasts.

D. What do influencers promote?

Finally, we wish to inspect the types of products being promoted by influencers. This can be done via our Instagram stories dataset, as each sponsored item is optionally tagged with the category of the advertiser. This is taken from a control set of tags offered by Instagram. Note that this is *only* available for stories, and not regular posts. We find that this feature is
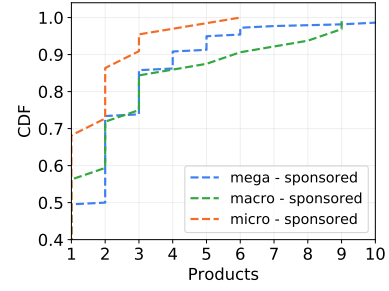

Fig. 5: CDF of number of products promoted by influencers across categories. This only covers stories because equivalent metadata is not available for posts.

not widely used by influencers, with only 3% stating their product.

Figure 5 presents a CDF showing the number of products promoted by influencers across categories. Most influencers only promote a single product, particularly in the case of micro. We observe that 50% *Mega*, 58% *Macro* and 70% *Micro* influencers promote just a single product. This suggests that influencers tend to focus on a particular product type, likely in their own specialist area.

We also inspect the type of products influencers use stories to promote. This metadata is captured by the Instagram API, although it only covers stories, as posts do not contain this explicit metadata. Figure 6 presents the top 20 product types influencers promote. The Y-axis counts the number of unique accounts promoting each type of product. We observe that most publishers tend to advertise products under *Health/Beauty* (14%), *Product/Service* (11%) and *Clothing (Brand)* (11%). Across products we observe that Mega influencers tend to dominate: they are the major publishers for 77% of product types we identify. This is in sharp contrast to Macro (14%) and for Micro (9%). These findings confirm the intuition that Instagram is dominated by promotions surrounding consumables such as food, retail and beauty. These cover 43% of all adverts.
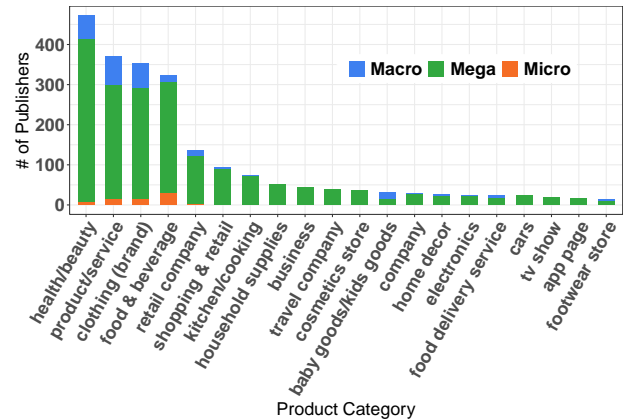

Fig. 6: Number of products promoted in stories based on their type (identified via the Instagram API).

V. PREDICTING SPONSORED CONTENT

The previous results have shown that sponsored posts tend to have unique properties that differentiate them from non-

sponsored content. This leads us to inspect the viability of automatically detecting (and tagging such posts). This, for example, could be used to protect users from being misled.

A. Classifier Design

We take a supervised deep learning approach, as we have a substantial ground-truth (annotated) dataset of sponsored *vs.* non-sponsored posts.

Dataset. For classification, we use the post dataset described in Section III-A. In total, the dataset consists of nearly 7K sponsored and 27K non-sponsored posts from all groups of influencers. Note that we do not differentiate between the different popularity groups and, instead, treat all accounts equally. As we are dealing with an imbalanced dataset, we use Random Under-Sampling to reduce the size of the non-sponsored class. Accordingly, we randomly select posts from non-sponsored class and remove them from the training dataset. While selecting the posts, we try to keep the diversity of influencer's profiles. So, we select random samples from influencers who have more published post. This process started from nano (with a higher number of post) to mega.

Finally, we produce a dataset with 14k observations with an equal number of sponsored and non-sponsored labelled posts (Section III-A). Note that all these posts are generated by users who have posted at least one sponsored post. Hence, our classification task is limited to differentiating between sponsored and non-sponsored posts from influencers (rather than a more general audience). Based on our observations, the posts in the above dataset can be separated into three sub-populations:

1) **Sponsored Post.** In a sponsored post, influencers normally try to directly or indirectly advertise a product. By adding sponsored metadata (Section III-A), the post is explicitly declared as an advertisement. It may also contain related hashtags for that product and the page of that product. We note that the main company is often tagged in the post.

2) **Non-Sponsored Post.** In contrast to a Sponsored Post, this is a routine post by an influencer which does not include any direct or indirect advertisements. As a result, the photo/video, caption, and list of hashtags are not organized or written in such a way to promote anything.

3) **Hidden Sponsored Post.** This sub-population is similar to the first item (Sponsored Post) except that sponsored metadata (*i.e.,* hashtags) are removed. We remind the reader that The Advertising Standards Authority (ASA) [21] advises the use of such hashtags, as any paid promotions are subject to consumer protection law. Despite this, we posit that there is an incentive for influencers to avoid disclosure, as there is greater value in "personal" endorsements. "*Hidden Sponsored Post*", "*Hidden Advertisement*" and "*Non-declared Sponsored Post*" terms address the same meaning in this paper.

Feature Engineering & Pre-processing. The next step is to build a set of features from the post content to train our model.

Fig. 7: An Example of non-declared sponsored post: 1) A verified Mega influencer holding a box on her hand and promoting products. 2) The product page is tagged in the photo and mentioned in the caption. 3) The caption invites us to use this product, but there is no sign of sponsored metadata. 4) The post received 4K likes and 500 comments.

Table III summarizes the features that are employed. We split the feature list into two main categories: post features and publisher features. Post features comprise all features that are obtained from the content of the post, *e.g.* number of likes, the caption. Publisher features cover features extracted from the profile of the publisher. For all text-based features such as "post caption" and "post hashtag", we remove all punctuation marks, stopwords and convert them to lowercase characters. Words are stemmed to reduce to their root forms.

TABLE III: Feature Set Used for Classification.

Post Features		Publisher Features	
Feature	**Type**	**Feature**	**Type**
post caption	*text*	number of follower	*numeric*
post hashtag	*text*	number of followee	*numeric*
number of likes	*numeric*	length of biography	*numeric*
number of comments	*numeric*	profile biography	*text*
length of caption	*numeric*	is verified	*numeric*
number of hashtags	*numeric*	external URL exist	*numeric*
number of mentions	*numeric*		
number of tagged users	*numeric*		

Model Architecture & Performance. Next, we propose a Contextual LSTM Neural Network architecture (presented in Figure 8) and we compare the result with Random Forest as a classic machine learning based Classifier. Post metadata can enrich the information available for classification. So, to build the model, we combine texts with other metadata features (Table III). First, we tokenize text metrics (*e.g.* caption and hashtag) using Keras Tokenizer Class [22] and then, the result is fed to the LSTM layer which outputs a 64-dimension vector. Then, we attach numerical metadata (post and profile features) to this vector and pass it through 2 ReLU activated layers of size 128 and 64. Finally, it connects to an output layer that predicts the label.

We use TensorFlow, Keras, and Scikit-learn libraries, and we run them on Google CoLab. Also, we use a random split of 80% (training set) and 20% (test set) and to avoid overfitting, we use 10-fold Cross-validation. For each model, we compute
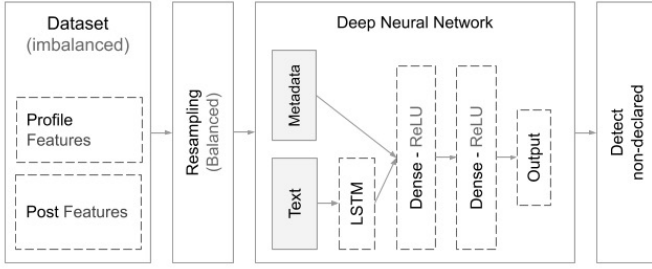
Fig. 8: General architecture of sponsored content detection (Contextual LSTM model).

TABLE IV: Classifier Performance

Model	Accuracy	Precision	Recall	F1
Random Forest	0.84	0.83	0.84	0.83
Contextual LSTM	0.89	0.88	0.87	0.89

the accuracy, precision, recall and F1 in Table IV. For our labelled data, we observe high performance, with positive results across both classifiers. Our Contextual LSTM Classifier enhances the performance by nearly 5% (89% of accuracy).

B. Results

We now evaluate our classifier across the entire dataset to quantify the prevalence of hidden sponsorship.

Detection. As we believe some influencers do not declare sponsored posts with defined metadata (§III-A), we use the DNN trained model on the non-sponsored posts which are not used in the training. This part includes nearly 20K posts from all 4 groups of influencers (mega to nano).
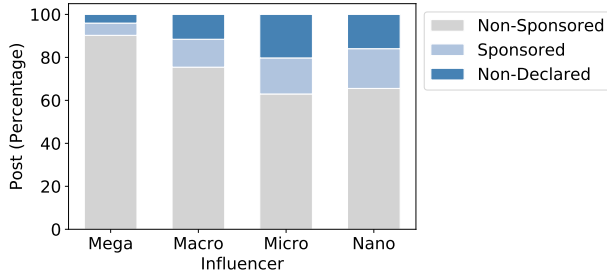

Fig. 9: Overall view of different post distributed by influencers.

Results. Our model detects that 17.7% sponsored posts are published without metadata (or not declared as sponsored) which is approximately 3.8K posts through all influencers. Figure 9 demonstrates the portion of *non-sponsored*, *sponsored* and *non-declared sponsored* post across each group. We see a noticeable set of hidden sponsored posts, particularly for Micro and Nano influencers. Nano with 18%, micro with 21%, and macro with 11% of shared posts, advertise almost 4 times more than Mega. In detail, by considering clusters of influencers we observe:

- **Mega.** Mega is the smallest group in terms of having hidden advertisements. Nearly 29% of Mega influencers publish one non-declared post. In detail, out of 3K non-sponsored posts, 4.3% are detected as hidden advertisement entities.

- **Macro.** 57% of Macro influencers have one hidden sponsored post. In detail, 13% of 3.6K non-sponsored posts are recognised as hidden advertisements.

- **Micro.** Micro is the largest group in terms of publishing non-declared advertisements. In general, 78% of Micro influencers publish one sponsored post. In detail, 24% of the 5K non-sponsored content is detected as hidden advertisements.

- **Nano.** More than 70% of nano influencers publish one non-declared advertisement. In detail, within 9.7K non-sponsored posts, 20% are detected as hidden advertisements.

Validation. To validate the above results, we manually validate 50% of the detected posts to confirm they are hidden sponsored posts. Across the identified content, we witness in 81% of posts, influencers are confirmed as promoting products, 11% are incorrect, and in 8% we are not able to verify they are commercial or not. This confirms that our model *is* effective at identifying influencer posts.

C. Text Feature Analysis

Finally, we manually explore which textual features (Table III) are most prominent in our prediction task. We do this to gain an understanding of what characteristics are most common specifically for (hidden) sponsored posts. Thus, we manually check 50% of the predicted posts and for each metric, we inspect the unique distinctions between sponsored and non-sponsored posts. This allows us to compile a list of differences and then count how often these difference emerge. We summarise our findings as follows:

1) **Post Caption** holds the most valuable information of the post in the shape of the text (Fig 7). Often the text includes details of products, promotions, names, *etc.* and encourages users to perform an action. An action could be to buy something, follow a page, join a competition, install an application, use a specified hashtag, *etc.* We also notice some peculiar way of writing the text or using keywords as follows. If X represents a *product* or *producer*: (*i*) the caption includes X as raw text in 94% of the sponsored posts. (*ii*) Also X is mentioned (followed by @ sign and becomes blue) in 91% of the posts. (*iii*) The caption could contain particular sentences such as "*thank you X*", "*many thanks to X*", "*X from this page*", "*my top choice is X*", "*go and follow X*", *etc.* . This happens in 78% of the sponsored posts. (*iv*) We observe call-to-action keywords such as "*link in bio*", "*download it*", "*watch my story*", "*Use discount code*" and "*comment/like to win*", *etc.* in 53% of the posts.

2) **Post Hashtag** is the next leading text feature which generally includes valuable information about the product and its producer. Influencers may use one or more hashtags for the corresponding product (Figure 7). For example, if *#outfit* is the main hashtag representing the product, we also observe *#outfitday*, *#oufit*, *#bestoutfit*,

#trendyoutfit, *#outfitstyle*, *etc.* In 97% of the sponsored posts, the name of the product or producer (or both) is listed as a hashtag. To highlight some hashtags, influencers may put them in the text while describing the products, or create a separate list after the text (or in some cases in the first comment by the influencer). Hashtag count also helps differentiate sponsored content. Sponsored posts one to get hashtags trending, or (*ii*) to get more visibility in Instagram explorer, or (*iii*) to be easily found in search results, regularly have longer hashtag length:

3) ***Profile Biography*** is another feature that improves the accuracy of our classifier. In general, influencers usually put relevant information in their profile biography (which could be temporary): (*i*) In 63% of the biographies, information such as "sponsor info", "campaign details" or "promotion code" as raw text exist. (*ii*) In 54% of the profiles, influencers put sponsor "hashtag(s)" (which becomes blue) or mention "product/producer page" (followed by @ sign). (*iii*) In 34% of the profiles, there is a call-to-action phrase with special keywords such as "follow", "buy", "sale", "watch", "join", "check out", "promotion", "more info" *etc.* (*iv*) Profiles may include an External URL (in 21% of the profiles) or YouTube link (in 11% of the profiles), which redirects users to the main product webpage, producer website, or full youtube review video. Regularly influencers use shortener tools to monitor the number of people clicking on it. Biography length is also helpful as influencers who do promotion usually have (*i*) promotion codes, (*ii*) sponsor contact detail, and (*iii*) product-related hashtags on their biography (longer biography).

VI. CONCLUSION & FUTURE WORK

This paper has performed the first large-scale analysis of influencer behaviour on Instagram. Rather than solely discovering "celebrities", our methodology has exposed a large array of influencer types, including highly targeted nano influencers. We have further dissected the differing behaviours of these influencers, to find that even though Mega influencers garner significant attention, nano influencers are more successful in engaging their audiences with a larger fraction of followers interacting with their posts. We have further trained a DNN model to classify posts as sponsored, allowing us to identify seemingly sponsored posts that are not properly declared. As influencer income is taxable, we note that this may have financial ramifications that go beyond the issues of deceiving consumers.

This is just the first step in a wider research agenda. We have several specific lines of future work. First, we wish to expand our analysis across multiple platforms (*e.g.* Tik Tok, YouTube) and to gain a deeper understanding of the strategies employed by influencers. Second, we wish to revisit our classifier to improve performance. Although our current implementation obtains good results, we manually found 11% misclassifications. Consequently, we wish to build on our work to devise classifiers that can automatically identify influencers. This is particularly important for identifying "ghost" influencers who do not explicitly identify their adverts. We posit that this could also be used to support social media giants in identifying misbehaviour within the advertisement space. Whereas in this paper, we have focused on influencers themselves, we are also curious to better understand the dynamics of *audiences* as well as the *companies* that use influencers.

REFERENCES

[1] Luis V. Casaló, Carlos Flavián, and Sergio Ibáñez-Sánchez. Influencers on instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*, 2018.

[2] ASA. An influencer's guide to making clear that ads are ads, 2018.

[3] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[4] Amandeep Singh, Malka N. Halgamuge, and Beulah Moses. 5 - an analysis of demographic and behavior trends using social media: Facebook, twitter, and instagram. In Nilanjan Dey, Samarjeet Borah, Rosalina Babo, and Amira S. Ashour, editors, *Social Network Analytics*, pages 87 – 108. Academic Press, 2019.

[5] Thorsten Hennig-Thurau, Kevin P Gwinner, Gianfranco Walsh, and Dwayne D Gremler. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1):38–52, 2004.

[6] Duncan Brown and Nick Hayes. *Influencer marketing*. Routledge, 2008.

[7] Sofie Biaudet. Influencer marketing as a marketing tool: The process of creating an influencer marketing campaign on instagram. 2017.

[8] NL Ewers. # sponsored–influencer marketing on instagram: An analysis of the effects of sponsorship disclosure, product placement, type of influencer and their interplay on consumer responses. Master's thesis, University of Twente, 2017.

[9] Vaibhavi Nandagiri and Leena Philip. Impact of Influencers from Instagram and YouTube on their Followers. *International Journal of Multidisciplinary Research and Modern Education*, 2018.

[10] Morgan Glucksman. The rise of social media influencer marketing on lifestyle branding: A case study of lucie fink. *Elon Journal of Undergraduate Research in Communications*, 8(2):77–87, 2017.

[11] Karen Freberg, Kristin Graham, Karen McGaughey, and Laura A Freberg. Who are the social media influencers? a study of public perceptions of personality. *Public Relations Review*, 37(1):90–92, 2011.

[12] Kristy Sammis, Cat Lincoln, and Stefania Pomponi. *Influencer marketing for dummies*. John Wiley & Sons, 2015.

[13] Eva Lahuerta-Otero and Rebeca Cordero-Gutiérrez. Looking for the perfect tweet. the use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, 64:575–583, 2016.

[14] Seungbae Kim, Jinyoung Han, Seunghyun Yoo, and Mario Gerla. How are social influencers connected in instagram? In *International Conference on Social Informatics*, pages 257–264. Springer, 2017.

[15] US Federal Trade Commission. Us federal trade commission. https://www.ftc.gov/news-events/press-releases/2017/09/csgo-lotto-owners-settle-ftcs-first-ever-complaint-against.

[16] Instagram. Official api graph instagram. https://developers.facebook.com/docs/instagram-api, September 2019.

[17] Instagram. Instagram hashtag search. https://developers.facebook.com/docs/instagram-api/guides/hashtag-search, September 2019.

[18] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P Gummadi. Measuring user influence in twitter: The million follower fallacy. In *fourth international AAAI conference on weblogs and social media*, 2010.

[19] Instagram. What are the requirements to apply for a verified badge? https://help.instagram.com/312685272613322, September 2019.

[20] Json Hjh. Instagram stories vs feed ads. which is more effective at driving traffic? https://www.agorapulse.com/social-media-lab/instagram-stories-ads, February 2018.

[21] The Advertising Standards Authority Ltd. (trading as ASA). The advertising standards authority. https://www.asa.org.uk.

[22] François Chollet et al. Keras. https://keras.io, 2015.