CROWDTURFING ON INSTAGRAM – THE INFLUENCE OF PROFILE CHARACTERISTICS ON THE ENGAGEMENT OF OTHERS

Research paper

Voronin, Georg, Humboldt University of Berlin, Berlin, Germany, voroning@outlook.com

Baumann, Annika, Humboldt University of Berlin, Berlin, Germany, annika.baumann@hu-berlin.de

Lessmann, Stefan, Humboldt University of Berlin, Berlin, Germany, stefan.lessmann@hu-berlin.de

Abstract

Crowdsourcing is the structured cooperation of a large number of individuals to accomplish a specific goal. Benefits of crowdsourcing are manifold but there is also risk to misuse corresponding services for illegal or unethical behaviour. Crowdturfing is a special form of crowdsourcing and encompasses actions aimed at manipulating information on social media websites. Being a relatively new concept, crowdturfing has not yet received much attention in the literature. The paper investigates the effect of crowdsourcing activity on the social media platform Instagram. Using an experimental design, we set up three different Instagram profiles with varying degrees of popularity. We govern the popularity of a profile by means of crowdturfing. The paper studies how the number of followers, likes and comments influences the engagement of third-party individuals. Our results show that profiles with greater popularity do have it easier to gain more attention from others in the form of more followings. Furthermore, we are able to show that Instagram is not robust towards crowdturfing activities as the popularity of content can be pushed through such non-organic activities.

Keywords: Crowdsourcing, Crowdturfing, Instagram.

1 Introduction

Social media influences everyday life with several websites reporting a large number of users, many of whom visit social media websites on a daily basis. Having such a large user base, there is a great potential for individuals and businesses to gain attention and influence over a large audience by means of this social network. The stronger the popularity signal on a profile page in terms of a large number of followers (Westerman et al., 2012; Omilion-Hodges and Rodriguez, 2014; Lee et al., 2018), reposts (Gao et al., 2015; Lin et al., 2016) and comments (Omilion-Hodges and Rodriguez, 2014) the more positive profile visitors perceive the credibility of content on social media. This enables content created on social media to have a severe impact on several aspects of daily life. For e-commerce applications a positive relationship between the perceived legitimacy of a social media profile and the purchase intention towards the product has been shown to exist (Lee et al., 2018). Furthermore, the impact of social media content is based on the early popularity of the content (Szabó and Huberman, 2010), which could be also utilized for harmful activities such as distributing fake news as happened in case of the US presidential election campaigns in 2016 (Allcott and Gentzkow, 2017).

A recent social media platform is Instagram, which accumulates a daily visitor count of around 500 million users (Systrom, 2017). The network has grown to around 80 million daily uploaded photos and videos (Instagram Inc., 2016a) causing a severe information overload. The average user misses 70 percent of the content on their feed (Instagram Inc., 2016b). Therefore, receiving popularity by others in the form of likes, comments and followings is a scarce commodity and to be able to spread content through the network and to receive more visibility across the borders of one's own social circle a certain popularity of a profile might be crucial. Based on a sample of around 21,000 Instagram profiles it has been shown that a large share of Instagram users follow only a small number of other profiles (Moore, 2011) making it challenging to receive attention, and thereby popularity, of novel users that have never interacted with ones profile before.

Such an environment might motivate the use of crowdturfing (CT). The term CT stems from a combination of crowdsourcing and astroturfing. Crowdsourcing is the structured collaboration of a large number of individuals towards a common goal (Martin et al., 2008). Astroturfing describes actions that seem to be natural and spontaneous but have a commercial or political motivation and are centrally controlled by a specific agency (Wang et al., 2012). Thus, CT is used to influence individuals on the basis of purchased activity in terms of likes, comments and followings generated by a large number of workers within a social network. First investigations show that CT is a widespread phenomenon. Motoyama et al. (2011) identify 30% of activities on crowdsourcing platforms as CT in 2011. For 2012, this amount has been estimated to be 95% (Wang et al., 2012). CT services target a variety of platforms such as Twitter or Google's search engine, but also social networks such as Facebook and Instagram (Lee et al., 2015; Choi et al., 2016; Rinta-Kahila and Soliman, 2017). Specific CT activities include the creation of profiles and content but also the manipulation of likes, followers, comments and re-sharing of content (Motoyama et al., 2011; Wang et al., 2012; Lee et al., 2013; Lee et al., 2015; Choi et al., 2016).

Social media has the ability to influence the opinion of its users, while at the same time information on those platforms is prone to manipulation through CT. CT is a rather recent phenomenon and several dimensions of it are not yet fully understood. The effects of CT activities on a visual content-based platform such as Instagram has not been studied before. Therefore, our paper aims to investigate the effect of CT activity on Instagram.

Our paper is able to make three contributions. First, we analyse the profile characteristics of social media accounts on Instagram associated with CT activity in comparison to organically developed profiles. We are able to show that significant differences do exist between those profiles in terms of number of followers, likes and other profile characteristics. Second, we aim to understand how profile characteristics stemming from CT activity on Instagram affect the engagement level of other users visiting a profile. To that end, we set up an experiment by creating three Instagram profiles which differ in their degree of activity and recognition by others which we control using CT activity. Posting content in a pre-defined frequency on each profile, we aim to receive attention of others by following their Instagram account. We measure whether they react by also following our account in response. We are able to show that the pre-existing popularity of an Instagram profile in the form of follower, like and comment count does have an influence on whether users are more likely to follow an Instagram account. Third, we are able to show that Instagram is vulnerable towards CT activities by flagging posts as being popular despite its artificially created attention through CT.

The structure of our paper is as follows: first, we summarize related work. Next, we describe our experimental setting and continue to presents the results. Last, we discuss and summarize our findings.

2 Background and Related Work

In general, four different players engage in CT activities (Wang et al., 2012). The first player are the customers paying for such a service, striving to generate more recognition concerning the CT campaign of interest. The second one are the workers who are the actual individuals performing the CT activity such as generating likes, comments or profile followings. The third player are the CT agents who are responsible for the distribution of work and the management of a campaign. They provide a platform for customers and workers to engage with CT activities. The last player are the social media platforms and their users, which the CT activity targets. Figure 1 illustrates the interaction between the four players in the CT circle. First, a customer orders a specific campaign on a CT platform, which is provided and organized by the agent. The customer has to pay for the service in the form of monetary value or digital coins earned through one's own CT activity. Workers on the CT platform can then decide to engage in this specific campaign through social media activity, for example in the form of likes, reposts, comments or followings. Workers receive an incentive for their actions in the form of money or digital coins, which they can exchange for CT activity in return. The CT activity of workers helps the customer to gain attention and popularity for the campaign of interest.

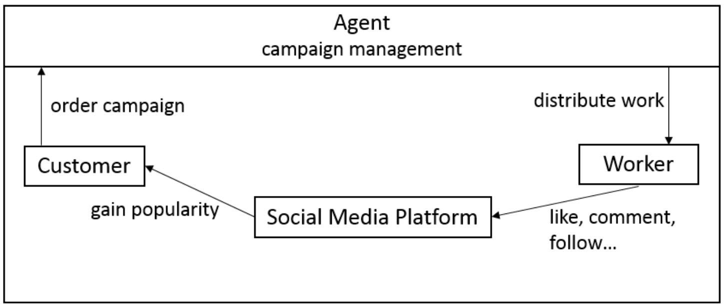

Figure 1. The CT circle (based on Wang et al., 2012).

A social media profile has the potential to deliver information, so-called cues, about the profile owner. In general, the Brunswik lens model (Brunswik, 1956) states that impressions of other individuals are formed by the various cues an individual provides. Those cues can be used to infer certain characteristics of a person and can be present both in the offline and online environment (Vazire and Gosling, 2004). In case of an online setting, such cues might be based on the characteristics a social media profile provides, for example system-generated features like the number of followers, comments and likes (Tong et al., 2008). According to the Brunswik lens model, such cues and therefore the aligned popularity of a profile and its content can be influential in terms of how others perceive and respond to a user's account. Several studies did indeed observe a positive effect on the perceived legitimacy and credibility of social media content based upon specific profile features like the number of followers (Westerman et al., 2012; Omilion-Hodges and Rodriguez, 2014), comments (Omilion-Hodges and Rodriguez, 2014) and how often content has been shared (Gao et al., 2015; Lin et al., 2016). However, also contrary effects have been observed where a large number of friends has the impact to raise certain doubts towards a profile in terms of the profile owner's social attractiveness (Tong et al., 2008).

Prior work examines CT on social media at different levels (see Table 1). We classify relevant studies with respect to the target platform and the player on which a study concentrates. Further, we classify prior research according to the main research questions, which we categorize as *Descriptive*, *Detection* and *Outcome*. *Descriptive* studies focus on high-level statistics such as what kind of CT activities are common (e.g., Wang et al., 2012; Lee et al., 2015) or which social media platforms are targeted (Choi et al., 2016). *Detection* studies focus on the automated recognition of CT content and activity (e.g., Song et al., 2015). *Outcome* studies investigate the effects of CT activities on social media platforms such as understanding the influence of CT activities on other individuals and the social media platform (Wang et al., 2012; Lee et al., 2015). Especially the algorithmic-based identification of workers, content and activity created through CT activity has been in the focus of research (17 out of 18). Furthermore, seeking insight into CT through descriptive statistics is common (11). Understanding the influence of CT activity on social media platforms has received relatively little attention so far (3).

Some papers investigate the outcomes of manipulating activity on online social networks. Wang et al. (2012) compare the response in terms of clicks on advertising campaigns that were spread through three types of social networks by the help of CT workers. Their results suggest that marketing initiatives in QQ instant message groups were most effective (in terms of clicks), followed by Sina Weibo. The authors could not verify an effect in case of forums. Lee et al. (2015) measure the impact of subscribers gained through CT on the Klout score of a profile. The Klout score measures the quantitative and qualitative influence of a social media profile by analysing the activities of the profile owner with other users (Klout Inc., 2015). To achieve this, Lee et al. (2015) equipped five Twitter profiles with a varying number of followers ranging from 600 to 32,000 subscribers. A larger number of followers always resulted in a higher Klout scores even if the popularity has been achieved through CT. Kim et al. (2016) investigate popularity effects of CT activity on Fiverr and conclude that such malicious activities only have a low impact on the popularity of their promoted content due to the closeness of the topological network of CT workers.

Reference	**Platform**	**Main Focus**	**Descriptive**	**Detection**	**Outcome**
Aggarwal 2016	Twitter	Agent	X	X	
Aggarwal et al. 2018	Twitter	Customer	X	X	
Aggarwal and Kumaraguru 2015	Twitter	Agent	X	X	
Badri Satya et al. 2016	Facebook	Agent, Worker	X	X	
Castellini et al. 2017	Twitter	Worker	X	X	
Choi et al. 2016	CT platforms, Facebook	Agent, Worker	X	X	
Cresci et al. 2015	Twitter	Worker		X	
Kim et al. 2016	Fiverr, Twitter	Worker	X		X
Lee et al. 2013	CT platforms, Twitter	Customer, Worker	X	X	
Lee et al. 2015	Fiverr, Twitter	Customer, Worker		X	X
Liu et al. 2016	Sina Weibo	Worker		X	
Shah et al. 2017	Twitter	Agent, Worker	X	X	
Shen et al. 2014	Sina Weibo	Customer, Worker		X	
Song et al. 2015	Twitter	Customer, Worker	X	X	
Stringhini et al. 2012	Twitter	Agent, Customer		X	
Wang et al, 2014	Sina Weibo	Worker		X	
Wang et al. 2012	Sina Weibo	Agent, Customer, Worker	X	X	X
Zhang and Lu 2016	Sina Weibo	Worker		X	

Table 1. Summary table of related work in the area of CT activities on social media platforms. We classify each work according to the main player(s), platform(s) and with respect to the main research question(s) it strives to answer.

Despite its growing popularity, Instagram has not been in the focus of research so far. By now, existing research analyses mainly the effects of CT on system-generated features (e.g., the amount of followers, likes and comments), or on how individuals perceive these features. But may the implications stemming from CT be more far-reaching than that? Does the distortion interact with the networks algorithms and is a user's manipulated perception reflected in his or her behaviour within the network? Both aspects were not studied before, but might be a real threat to the social network. Therefore, we strive to understand how CT influences the engagement of other users with the profile and how it interacts with parts of the Instagram algorithm. Pursuing this goal, we contribute to the literature in three ways. First, we provide insights on the existence and characteristics of CT profiles on Instagram. Second, we examine how profile features such as the number of followers, likes and comments influence how individuals interact with profiles on Instagram. Third, we show that CT activities indeed have an effect on some algorithms, raising the popularity of content to a significant extent.

3 Methodology

This section explains our methodological approach. First, we explain our process to identify profiles associated with CT activity and organically developed profiles. Next, we describe the experimental design to understand how initial popularity gained through CT of an Instagram profile influences the activity of other individuals and the effect on Instagram itself.

3.1 Characteristics of crowdturfing profiles on Instagram

In the first step, we aim to understand how the profiles of different players associated with CT activities differ from each other. To identify profiles associated with CT activity and those who are not, we register for a CT agency to gain access to worker and customer profiles. In the next step, we create two Instagram profiles. In order to be able to identify Instagram profiles of workers, i.e. individuals who perform CT actions, we request CT activity in the form of followings for our first profile from a CT agent. This allows us to identify 429 profiles that represent CT workers. To be able to identify profiles who represent CT customers, i.e. individuals who request CT activity, we use the second profile. We use this profile as our own worker profile to perform CT tasks (i.e., following other Instagram accounts). In total, we follow 429 Instagram profiles that requested CT activity from the agent therefore representing our identified customers.

To gain a deeper understanding of CT on Instagram, we want to compare the characteristics of the identified worker and customer profiles associated with CT to a sample of ordinary Instagram users who did not make use of such a service. Research shows that distinguishing between regular profiles and profiles associated with CT is challenging (Song et al., 2015). For example, CT customer profiles have been shown to be used by ordinary users and business profiles (Stringhini et al., 2012). Even celebrities and verified profiles may be associated with CT (Shen et al., 2014; Aggarwal and Kumaraguru, 2015). Therefore, it is difficult to conclude with certainty that a profile has not undergone some sort of CT. With respect to worker profiles, some automated detection approaches are available for Twitter (e.g., Aggarwal and Kumaraguru, 2015), Facebook (Badri Satya et al., 2016) or Sina Weibo (e.g., Wang et al., 2012). To the best of our knowledge, an automated identification of worker profiles on Instagram has not been considered. However, we assume that Instagram itself does not want to be associated with such harmful activities and has strong incentives to prevent an undue popularity increase from CT of its own profile. We further assume that the official Instagram profile has no followers that are associated with CT. On that basis, we collect 429 profiles of the list of followers of the official Instagram profile which we assume to be organic, i.e. not making use of CT activity.

We collect data from these 1,287 profiles (429 CT workers, 429 CT customers, 429 ordinary users) to analyse several dimensions of characteristics for each profile group. We collect information such as privacy settings and further profile information (existence of profile photo, name, website, and biography). In addition, we acquire information about whether they have any content published, and whether they constitute a business profile or a verified profile.

3.2 The influence of crowdturfing on third-party users

In the second step, we aim to understand what effect popularity gained through CT activity has on the reaction of other Instagram users. To do so, we set up an experiment using three almost identical Instagram profiles. The profiles have the same name, profile picture, biography, and follow the 50 most subscribed Instagram profiles in 2016 according to Social Blade LLC (2017). The Instagram usernames of the profiles were chosen by using an online random generator for names. Furthermore, all three profiles have the same ten pictures uploaded, which serve as initial content. While an identical profile picture is posted on each account, the content itself differ by the depicted objects to make the respective gallery look organic. All pictures stem from the same source, are unedited, have been taken by us with the same device and do belong to the category of architecture. More precisely we decided to use buildings as visual content, since we wanted to reduce a potential bias as possibly present for instance in case of pictures showing faces or animals (Bakhshi et al., 2014).

	No Popularity Profile	Moderate Popularity Profile	High Popularity Profile
Follower	0	100	10,000
Likes per Post	0	7	750
Comments per Post	0	1	55

Table 2. Characteristics of the three Instagram profiles for the experimental setting.

All three profiles differ in certain characteristics to simulate different popularity levels (Table 2). One profile resembles a profile having no popularity. It comprises the above activity but no other popularity indicators such as having followers, likes or comments. The profile with moderate activity has 100 followers and each individual picture has a total of seven likes and one comment. The last profile represents a popular profile. It has 10,000 followers and each picture accounts for 750 likes and 55 comments. The resultant profiles are shown in Figure 2.

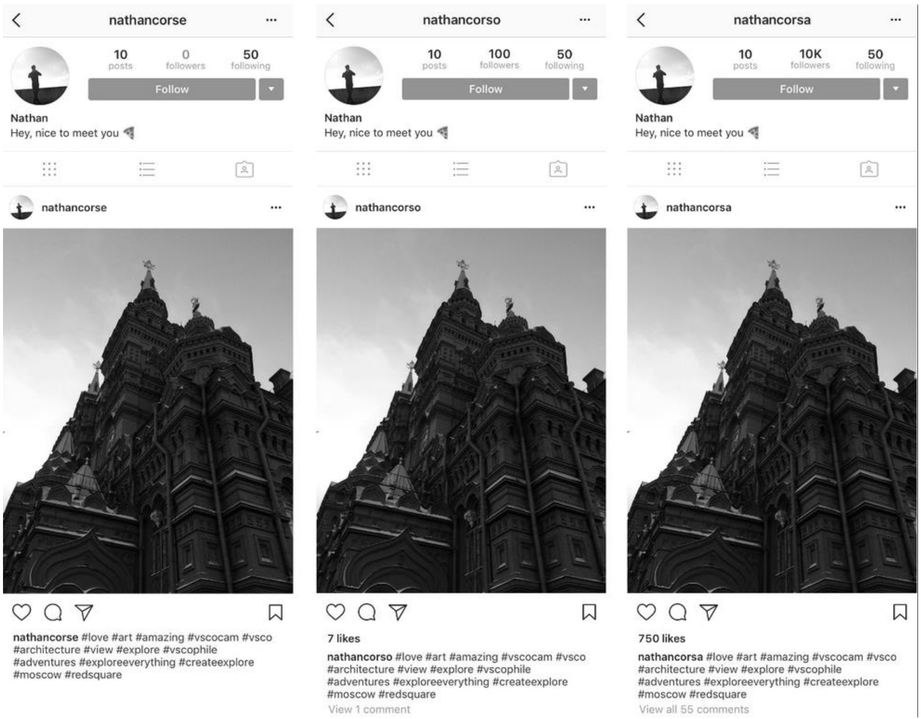

Figure 2. Screenshots of the three profiles showing the profile header and an example for content uploaded over the course of the experiment.

For each of those profiles and posts as well as all content uploaded over the course of the experiment, we make use of CT application to ensure the above mentioned popularity. After constructing the three Instagram profiles, we collect data over a period of 32 days. On each day, we post simultaneously one new photo on each profile and add the CT-based popularity to it. Then, we draw the attention of ordinary users towards our profiles by subscribing to their profiles and analysing their responses; if any. Afterwards, we reset profiles to their initial stage. This way, each day can be considered an iteration in a repetitive trial. Such an iterative setup helps to reduce variance and increases the robustness of observed results. Figure 3 depicts the setup. Subsequent sections detail the experimental procedures performed for each profile and day.

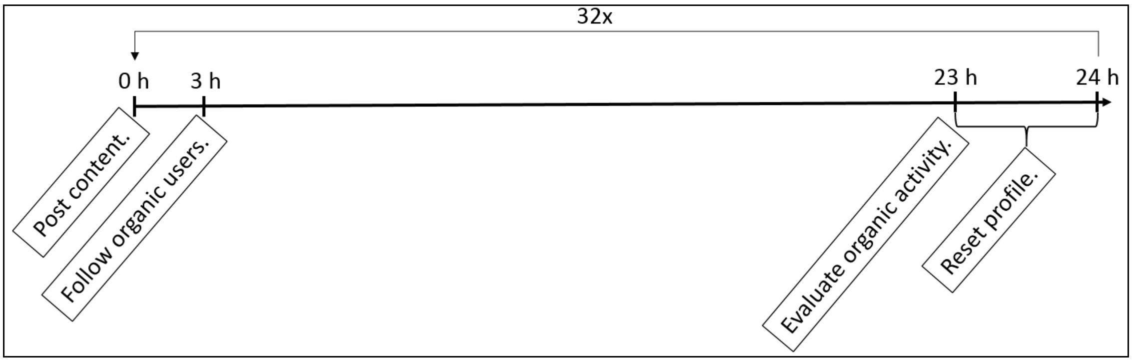

Figure 3. General setup of our experiment spanning a period of 32 days.

3.2.1 Step 1: Creation of content and crowdturfing-based popularity

In the first stage, we publish the same novel content on each of our own three Instagram profiles. Recall that we consider content showing some architectural structure; typically a building. We also delete the first posted picture on each profile, to keep the number of content per profile fixed at ten. Once the new content is available online, we add the relevant amount of likes and comments to each profile which we obtain through CT. The profile having no popularity does not receive any comments, likes or followers. The profile with a moderate popularity, having 100 followers, does receive seven likes and one comment for the newly uploaded picture. The most popular profile, the one holding 10,000 followers, receives 750 likes and 55 comments for the new uploaded picture. Since activity obtained through CT services might take some time to accumulate to the requested number, we wait for a period of three hours until all comments, likes and followings have been received on each profile. This way, we are able to separate the activity resulting from CT and the following steps.

3.2.2 Step 2: Raising organic popularity

In step two, each of our three profiles follows a total of 200 ordinary users to trigger a potential reaction from them using the subscription function of Instagram. Again, based on the assumption that Instagram itself does not want to contribute on damaging their own ecosystem, we select the user from the follower list of the official Instagram page. Each user was chosen at random to minimize a potential bias through pre-selection. It is important to note that each user is followed only once over the course of the whole experiment period to not cause any bias through repeated followings. After a user has been followed, he/she might decide to look at our profile. The respective number of followers was visible at the top of each of our Instagram profiles and the manipulated number of likes and comments for each picture could be observed upon viewing the content. The users we followed could then decide whether they respond to our profiles' subscription by following us back or by liking and/or commenting on one of our posts. Since nearly two-thirds of Instagram users visit the platform daily (Systrom, 2017), we decided to wait for a period of 20 hours for a user to visit his/her Instagram account to discover our following activity and then to decide whether to take action or not.

3.2.3 Step 3: Measuring organic popularity

After 20 hours have passed, we count the number of relevant variables on each profile which are X_1: the total amount of gained followers, X_2: the total amount of gained likes, and X_3: the total amount of gained comments. In case of public Instagram profiles we were easily able to correctly trace back the received actions since they show up in our list of followers on our Instagram profiles therefore enabling us to separate activity coming from one of the 200 participants we followed during step 2 of our experiment or from other sources. For private profiles, we used a list of temporary saved usernames to be able to adequately distinguish their activity from activity resulting from external factors. By doing so, we are able to identify activity obtained through our own following actions.

3.2.4 Step 4: Establishing the original state

In the final step, we restore each profile's original condition by deleting all accumulated activity we received during the 20 hours period. We delete subscriptions made at stage 2 of our experiment by making use of the function of blocking and unblocking a user, which makes him unfollow our profile automatically. Furthermore, we now check that each profile still has the number of followers obtained from CT at the beginning of the experiment. In case deviations were present, we restored the initial state on each profile by using CT activity. This was done to ensure stable condition for each iteration of our experiment.

3.3 Popularity of crowdturfing-boosted content

In the third part of our study, we strive to understand how CT affect content popularity on Instagram. For this purpose we examine if the content from our artificial profiles published at step 1 can make their way in the top posts of Instagram's hashtag and place pages. This is a section where trending posts are placed. Contrary to our first experiment, it is not possible for us to trace back any kind of reaction of users stemming from this source since we do not know if possible actions come from a post being placed in the top posts section or from another source like for example the personal Instagram feed, where users get shown posts based on their interests.

To assess whether the mechanisms of Instagram are able to filter out deliberately popularized content, we add 16 tags to every new post we publish at step 1 on each profile. We include a total of 13 hashtags that are always the same over the course of the experiment. Those hashtags are "#love", "#art", "#amazing", "#vscocam", "#vsco", "#architecture", "#view", "#explore", "#vscophile", "#adventures", "#exploreeverything", "#createexplore", and "#hype". We chose to use very popular hashtags with an average activity level of 903.59 likes and 11.61 comments for the top posts, whose usage spanning from almost one billion to slightly more than two million tagged posts. This was done since popular posts only appear for trending hashtags (Instagram Inc., n.d.). In addition, we use three dynamic tags comprising two hashtags and one geotag describing the post's location by mentioning the name of the corresponding city and country. After finishing step 1 of our experiment, i.e. gaining popularity through CT for each post, we wait for five minutes, before we analyse all 16 top posts sections for our artificial profile's posts. In the end of the 32 day spanning experiment, we count the number of postings for each profile that have made their way into at least one top post section.

3.4 Ethical considerations

The experimental design involves engagement in CT activity. This may raise concerns related to compliance with terms and conditions as well as community standards. Over the course of our research we payed close attention to account for ethical and legal considerations based on the terms of use of Instagram (Instagram Inc., 2017) and we followed several precautionary measures as proposed by other papers, which conduct research towards underground markets (Thomas et al., 2013; Song et al., 2015). We will further highlight the precautions taken in the subsequent paragraphs.

Our experiment did involve users of Instagram. Due to the nature of our experiment it was not reasonably possible to request informed consent. The analysis of social media data and user behaviour on social media platforms without explicit consent has been a common approach in research to eliminate the Hawthorne effect (e.g. Quercia et al., 2011; Hanson et al., 2013). The Hawthorne effect states that knowledge of being part of an experiment might lead to unwanted behavioural changes of participants (McCarney et al., 2007). We argue that our actions did not adversely affect users because our artificial accounts, even though created for research purposes, did provide relevant content for the platform and its users. In addition, we only share content on our profiles at an interval of 24 hours, therefore constituting no harmful or spam-like activity. We do not analyse specific content of social media profiles but instead focus on publicly available profile characteristics and the observation of behaviour related to one of the main functionalities of social media platforms, i.e. the subscription to one's account. The users we followed could decide themselves whether they want to react or not. We made sure that users were followed at maximum only once over the course of the whole experiment. Furthermore, we applied no further interactions that may be unwanted or being considered as spam. Therefore we minimized any potential risks for each individual involved. At the end of our experiment, all profiles and activities created were immediately deleted from the platform to cause no further distortions.

Data collection was performed manually and involved only publically available information. All users involved in the experiment remained anonymous and we applied several mechanism to ensure privacy of all participants involved such as only using aggregated data.

The use of CT services also raises some ethical concerns. In this regard, we engaged in CT activity entirely for the purpose of conducting research, aimed at minimizing any type of risk towards Instagram and its users. To analyse the effect of CT activity on Instagram, it was mandatory to make use of such services even though they are not desirable from an ethical perspective. However, we suggest that research in this regard is necessary to further understand the nature of CT and to potentially define countermeasures against such harmful activities. Similar studies related to the analysis of CT activity followed a similar approach (e.g. Thomas et al., 2013; Song et al., 2015).

4 Results

The section provides our empirical results concerning the characteristics of profiles associated with CT activity in comparison to organic profiles, and how CT affects popularity on Instagram.

4.1 Characteristics of crowdturfing profiles on Instagram

Table 3 presents characteristics for three groups of Instagram profiles. The comparison shows how profiles of organic users in our sample differ from those of CT customers and workers.

	Organic User	**CT Customer**	**CT Worker**
Number of profiles	429	429	429
Private	41.72%	10.96%	15.38%
Posts published	90.14%	96.50%	47.79%
Further profile information presented	99.77%	99.77%	87.41%
Profile picture	98.14%	98.60%	54.78%
Name	93.71%	97.67%	78.55%
Website	11.89%	23.31%	5.36%
Biography	59.91%	88.81%	30.77%
Business profile	1.63%	12.82%	2.33%
Verified profile	0.00%	2.33%	0.00%

Table 3. Summary of profile features regarding profiles associated with CT (customers and workers) and organic Instagram users.

The most striking difference concerns the fraction of private profiles. Whereas more than forty percent of the organic profiles are private, the faction of private profiles among CT customer and worker profile is only 10.96 % and 15.38 % respectively. Considering the amount of business profiles reveals further dissimilarities. The sample of organic and worker profiles account for only 1.63% and 2.33% of business profiles. Among CT customers, 12.82% of profiles relate to business organizations. Also, nearly 100% of organic users and customers reveal further information on their profiles whereas only 87.4% of workers do so. Furthermore, only 98.14 % and 98.60 % of organic users and customers, respectively, present a profile photo. The majority of those groups also have a name and at least one post published. However, only 54.78 % of the 429 worker profiles show a profile photo, whereas nearly half of them have posted something.

Looking at a more detailed analysis of profile characteristics with respect to number of posts, followers and followings more dissimilarities among the three groups become visible (Table 4). Based on our sample, customer profiles seem to be the most active ones having around 335 posts on average. However, the large standard deviations signals that customer profiles differ to a large extent in terms of the number of posts. Organic users are still fairly active accumulating around 140 posts on average, however with a large gap compared to customer profiles. Worker profiles seem to be barely active in terms of publishing content, having only around 17 posts uploaded on average.

Not surprising is that profiles associated with CT are particularly active when it comes to the number of followers and followings. Customer profiles have on average more than 45,000 follower but only around 880 individuals who follow them. Comparing the number of followers and followings, only 14.45% of customers' profiles have less followers than followings. In contrast, organic users constitute the profiles with the lowest number of followers and followings. Worker profiles are the only group who have on average more followings than follower. Around 90.21% of those profiles do have less followers than followings. Profiles associated with CT have far more subscribers than organic users, and partly because there are customer profiles with very few followings, resulting in extreme values for those groups.

	Organic User		**CT Customer**		**CT Worker**	
	Mean	StD	Mean	StD	Mean	StD
Posts	140.42	308.15	334.87	973.29	17.68	59.72
Follower	558.15	2,615.81	45,047.77	237,498.72	1,043.41	4,648.42
Followings	328.14	474.65	881.06	1,435.93	2,341.77	2,402.17

Table 4. Summary table of the characteristics, their means and standard deviations, of profiles associated with CT (customers and workers) and organic users.

4.2 The influence of crowdturfing on third-party users

The following section presents results on the relationship between the popularity of an Instagram profile and the engagement of third-party users with respect to whether they are more likely to i) follow a profile, ii) like a profile's posts or iii) comment on the profile's posts. Based on the results of our experimental set-up, Table 5 presents the means and standard deviations of the obtained data across the 32 days lasting research period.

Variables	**No Popularity**		**Moderate Popularity**		**High Popularity**	
	Mean	StD	Mean	StD	Mean	StD
Followers	23.53	6.08	23.97	3.87	26.56	4.41
Likes	44.78	15.56	47.56	17.01	47.03	14.54
Comments	0.72	1.9	0.59	1.29	0.78	1.66

Table 5. Means and standard deviations for the three conditions on the dependent variables.

We first analyse our data in terms of homoscedasticity and distribution via the Levine's and Shaprio–Wilk's Test. In case of X_3 (total amount of gained comments) the test indicates a violation of the assumption of a normal distribution on all three profiles. Therefore, we perform three one-way ANOVA (H_0: $\mu_{NoPopularity} = \mu_{ModeratePopularity} = \mu_{HighPopularity}$), one for each popularity measure, and an additional one-way ANOVA on ranks/Kruskal–Wallis test for X_3 to double-check the result of the first ANOVA. Our results show that while the popularity of a profile has a significant relationship with the engagement in terms of followers ($F[2, 93] = 3.61, p = .031, n^2 = .072$), neither the engagement in terms of likes ($F[2, 93] = 0.43, p = .755, n^2 = .006$) nor in terms of comments ($F[2, 93] = 0.11, p = .897, n^2 =.002$) can be supported. Our observations indicate significant differences among the no, moderate, and high popularity conditions on the dependent variable X_1 (number of followers). Next, we perform multiple pairwise tests using the Bonferroni method for each condition on X_1. This post hoc analysis demonstrates that users confronted with the high popularity profile were more encouraged to follow the profile than users confronted with the moderate popularity profile ($p = .015$) or the profile without any popularity ($p = .026$). However, there is no significant difference in the number of followers gained between the lowest and middle level ($p = .733$). Therefore, we may conclude that the data supports a positive relationship between the popularity of a profile and engagement in terms of following this profile. However, the observed results provide insufficient evidence to conclude that engagement in terms of liking a profile or commenting on content is significantly related to popularity of a profile.

4.3 Popularity of crowdturfing-boosted content

To assess how CT activity might be able to actually push the attention of content on Instagram, we examine the degree to which the uploaded pictures succeed in being mentioned in the top posts section of Instagram.

We observe that each post coming from the high-popularity profile was able to be placed in the top posts section for at least one hashtag used. We observe this result even in case when using the most popular hashtag #love which accounts to one billion posts at that time. In case of the profile with moderate popularity, we observe a top posts ranking for three uploads. This was the case for the tags #placareial, #wanchai, and #oia, whereas all of them refer to a specific geographic location. Therefore, we could only achieve to be placed in the top posts section by means of using geography-based tags which are not that popular, i.e. being tagged less than one million posts at that time.

Surprisingly, the content which received a mention in the top posts section in case of the moderate popularity profile was not able to do so in case of the high-popularity profile, despite accumulating an even higher popularity through CT. This might be a signal that Instagram provides certain countermeasures against deliberately popularized content which only is effective after exceeding a certain threshold. In case of the profile without popularity no placement in the top posts section of Instagram could be achieved which is not surprising since those posts did not receive any activity in form of comments or likes at all.

5 Discussion

To gain an understanding of CT on Instagram, we first analyse how profiles on Instagram associated with CT differ from those of ordinary users who do not engage in such an activity. There is a large proportion of private accounts among the organic Instagram users, which is more than twice as large as the share in the user groups that participate in CT. This may be due to the fact that owners of customer profiles are more interested in getting new subscribers and becoming popular. A private account is not suitable for such purpose. Furthermore, customer profiles account for the highest amount of Instagram business profiles and verified accounts. Therefore, our findings are in line with the results of Shen et al. (2014) and Aggarwal and Kumaraguru (2015). Furthermore, we observe a high number of worker profiles with a relatively low number of published content indicating low maintenance profiles. First, this might be a characteristic helpful to automatically identify such profiles. Second, such low maintenance profiles signal that a major fraction of worker profiles is engaged into CT activities just for financial reasons, earning real or digital money to spend it outside of the network or on CT activity for their main profile.

In the next step, we assess the effects of different popularity measures, namely the profiles' follower count and a posts' likes and comments, which we obtain through CT, on the behaviour of organic Instagram users. We are able to confirm the direct impact of the highest popularity on new subscribers. Individuals are more likely to follow an account when it seems to be popular compared to a profile with a lower level of popularity. Reason for this may be how a profile with so many followers is perceived by others, as research demonstrates that the sense of credibility of an information source is influenced by web-based metrics like the number of followers a profile has (e.g., Westerman et al., 2012; Lee et al., 2018). However, we suspect that there are more characteristics that are affected by those metrics, as a profile publishing images of architecture may not necessarily be measured in terms of credibility. This highlights the importance of implementing effective counter-measures towards CT. Profiles manipulating their popularity through CT additionally benefit from the advantage of an increasing commitment of other third-party individuals towards their profiles. Therefore, other profile owners might feel forced to engage into CT in order to restore balance. This is particularly important for businesses trying to gain influence in the network and may be the reason for the overrepresentation of business and verified profiles in CT activity. In the end this cycle might damage the social network of Instagram as a whole, its credibility and users.

The other variables, i.e. the number of comments and likes, do not show a statistically significant difference. This may be because not all activity levels affect each variable in the same way. As the number of comments and likes is only displayed upon viewing a post, more engagement of a user with a profile might be necessary. However, the results of those two popularity measures lacked empirical evidence to conclude on a significant relationship with other users' responses.

Last, our experiment shows that it is possible to place content of our profiles which have been popularized using CT-based likes and comments in the top posts section of Instagram. The algorithm responsible for selecting popular posts seems to lack robustness towards CT activity, which makes determining and eliminating CT on Instagram more important. It is also interesting to note that the content that came from the profile with the moderate popularity was in three cases preferred to content stemming from the high-popularity profile despite higher activity in terms of likes and comments. This might be a signal that Instagram already started to employ an algorithm to automatically detect deliberately popularized content.

6 Summary, Limitations and Future Outlook

In general, our results show that Instagram users who increase their popularity by manipulating the number of followers, likes, and comments to a higher popularity level, benefit from having a greater proportion of other individuals willing to follow them. The commitment of other users thus increases directly, while Instagram itself is not fully robust towards CT activities as the popularity of content can be pushed through such non-organic activities. This implies that specific counter-measures need to be developed further by Instagram to be able to adapt to and overcome deliberately manipulated content and activity.

One drawback of the study is that the activity gained was traced back through subscriptions and usernames. Thus, a private user may have changed his/her Instagram username during the period of the experiment so that actions based on the profile could not be assigned to the right profile, when they also denied our request for following them. Furthermore, the algorithm of Instagram which selects the top trending posts might change over time and therefore the results of our study might not be fully generalizable. In addition, our work is subject to the limitations of manual data collection. Accordingly, the variety of profile characteristics we were able to investigate is limited to the extent that we make use of only three artificial profiles. In case of the profile with the highest popularity, which is also the only one showing a significant impact on the received organic popularity, there exist a large difference in terms of the independent variables investigated. Thus, it makes it difficult to assess the actual threshold that enables the effect we search for. Therefore, future research could address this limitation using more fine-grained differences among the profiles to investigate. However, our research enables to derive first implications towards CT activity. Furthermore, investigating other scenarios like sharing different types of content should enable future research to gain more knowledge regarding the CT activity and its effects on Instagram. Furthermore, using machine-learning algorithms to automatically detect CT profiles on Instagram will be helpful to develop countermeasures against such harmful activities on the social network.

References

Aggarwal, A. (2016). Detecting and Mitigating the Effect of Manipulated Reputation on Online Social Networks. In: *Proceedings of the 25th International Conference Companion on World Wide Web*, 293–297.

Aggarwal, A. and P. Kumaraguru (2015). "What they do in shadows: Twitter underground follower market." In: *Proceedings of the 13th Annual Conference on Privacy, Security and Trust*, 93–100.

Aggarwal, A., Kumar, S., Bhargava, K. and P. Kumaraguru (2018). "The Follower Count Fallacy: Detecting Twitter Users with Manipulated Follower Count." In: *Proceedings of the 33rd ACM SIGSAC Conference on Applied Computing*.

Allcott, H. and M. Gentzkow (2017). "Social media and fake news in the 2016 election". *Journal of Economic Perspectives* 31 (2), 211–236.

Badri Satya, P. R., Lee, K., Lee, D., Tran, T. and J. J. Zhang (2016). "Uncovering Fake Likers in Online Social Networks." In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2365–2370.

Bakhshi, S., Shamma, D. A. and E. Gilbert (2014). "Faces engage us: Photos with faces attract more likes and comments on Instagram." In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 965–974.

Brunswik, E. (1956) "*Perception and the representative design of psychological experiments*. UnivUniversity of California Press, CA: Berkeley.

Castellini, J., Poggioni, V. and G. Sorbi (2017). "Fake Twitter followers detection by denoising autoencoder." In: *Proceedings of the International Conference on Web Intelligence*, 195–202.

Choi, H., Lee, K. and S. Webb (2016). "Detecting malicious campaigns in crowdsourcing platforms." In: *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 197–202.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A. and M. Tesconi (2015). "Fame for sale: efficient detection of fake Twitter followers." *Decision Support Systems* 80, 56–71.

Gao, Q., Tian, Y. and M. Tu (2015). "Exploring factors influencing Chinese user's perceived credibility of health and safety information on Weibo." *Computers in Human Behavior* 45, 21–31.

Hanson, C. L., Burton, S. H., Giraud-Carrier, C., West, J. H., Barnes, M. D. and B. Hansen (2013). "Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students." *Journal of Medical Internet Research* 15 (4).

Instagram Inc. (2016 a). *A New Look for Instagram*. URL: http://blog.instagram.com/post/144198429587/160511-a-new-look (visited on 11/24/2017).

Instagram Inc. (2016 b). *See the Moments You Care About First*. URL: http://blog.instagram.com/post/141107034797/160315-news (visited on 11/24/2017).

Instagram Inc. (2017). Terms of Use. URL: https://www.facebook.com/help/instagram/478745558852511https://www.facebook.com/help/instagram/478745558852511 (visited on 04/18/2018).

Instagram Inc. (n.d.). What are Top Posts on hashtag or place pages?. URL: https://help.instagram.com/701338620009494 (visited on 11/26/2017).

Kim, H. J., Chae, D. K., Kim, S. W. and J. Lee (2016). "Analyzing crowdsourced promotion effects in online social networks." In: Proceedings of the 31st Annual ACM Symposium on Applied Computing, 820–-823.

Klout Inc. (2015). *Der Klout Score*. URL: https://klout.com/corp/score (visited on 11/24/2017).

Lee, K., Tamilarasan, P. and J. Caverlee (2013). "Crowdturfers, Campaigns, and Social Media: Tracking and Revealing Crowdsourced Manipulation of Social Media." In: *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 331–340.

Lee, K., Webb, S. and H. Ge (2015). "Characterizing and automatically detecting crowdturfing in Fiverr and Twitter." *Social Network Analysis and Mining* 5 (1), 1–16.

Lee, Y. J., Yoon, H. J. and N. H. O'Donnell (2018). “The effects of information cues on perceived legitimacy of companies that promote corporate social responsibility initiatives on social networking sites.” *Journal of Business Research* 83, 202–214.

Lin, X., Spence, P. R. and K. A. Lachlan (2016). “Social media and credibility indicators: The effect of influence cues.” *Computers in Human Behavior* 63, 264–271.

Liu, Y., Liu, Y., Zhang, M. and S. Ma (2016). “Pay Me and I'll Follow You: Detection of Crowdturfing Following Activities in Microblog Environment.” In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence,* 3789–3796.

Martin, N., Lessmann, S. and S. Voß (2008). “Crowdsourcing: Systematisierung praktischer Ausprägungen und verwandter Konzepte.“ In: *Proceedings of the 2008 Multikonferenz Wirtschaftsinformatik*, 273–274.

McCarney, R., Warner, J., Iliffe, S. S., van Haselen, R., Griffin, M. and P. Fisher (2007). “The Hawthorne effect: a randomised, controlled trial.” *BMC Medical Research Methodology* 7 (30), doi:10.1186/1471-2288-7-30.

Moore, R. J. (2011). Instagram Now Adding 130,000 Users Per Week: An Analysis. URL: https://techcrunch.com/2011/03/10/instagram-adding-130000-users-per-week/ (visited on 11/26/2017).

Motoyama, M., McCoy, D., Levchenko, K., Savage, S. and G. M. Voelker (2011). “Dirty jobs: The role of freelance labor in web service abuse.” In: *Proceedings of the 20th USENIX Conference on Security,* 14–29.

Omilion-Hodges, L. M. and M. Rodriguez (2014). “The effect of system generated cues on users' perceptions of organizational trustworthiness, competence and goodwill.” *Journal of Media and Communication Studies* 6 (10), 161–164.

Quercia, D., Ellis, J., Capra, L. and J. Crowcroft (2011). “In the mood for being influential on twitter.” In: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom),* 307–314.

Rinta-Kahila, T. and W. Soliman (2017). “Understanding Crowdturfing: The Different Ethical Logics Behind the Clandestine Industry of Deception.” In: *Proceedings of the 25th European Conference on Information Systems,* 1934–1949.

Shah, N., Lamba, H., Beutel, A. and C. Faloutsos (2017). “OEC: Open-Ended Classification for Future-Proof Link-Fraud Detection.” *arXiv preprint arXiv:1704.01420*.

Shen, Y., Yu, J., Dong, K. and K. Nan (2014). “Automatic fake followers detection in chinese microblogging system.” In: *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 596–607.

Social Blade LLC (2016). *Top 50 Instagram Users by Followers*. URL: https://socialblade.com/instagram/top/50/followers (visited on 11/20/2016).

Song, J., Lee, S. and J. Kim (2015). “Crowdtarget: Target-based detection of crowdturfing in online social networks.” In: *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security,* 793–804.

Stringhini, G., Egele, M., Kruegel, C. and G. Vigna (2012). “Poultry markets: on the underground economy of twitter followers.” In: *Proceedings of the 2012 ACM Workshop on online social networks*, 1–6.

Systrom, K. (2017). *Strengthening Our Commitment to Safety and Kindness for 800 Million*. URL: http://blog.instagram.com/post/165759350412/170926-news (visited on 11/24/2017).

Szabó, G., and B. A. Huberman (2010). “Predicting the popularity of online content.” In: *Communications of the ACM* 53 (8), 80–88.

Thomas, K., McCoy, D., Grier, C., Kolcz, A. and V. Paxson (2013). “Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse.” In: *Proceedings of the 22nd USENIX Security Symposium,* 195–210.

Tong, S. T., Van Der Heide, B., Langwell, L. and J. B. Walther (2008). “Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook.” *Journal of Computer-Mediated Communication* 13 (3), 531–549.

Vazire, S. and S. D. Gosling (2004). "e-Perceptions: Personality impressions based on personal websites." *Journal of Personality and Social Psychology* 87 (1), 123–132.

Wang, G., Wang, T., Zheng, H. and B. Y. Zhao (2014). "Man vs. machine: Practical adversarial detection of malicious crowdsourcing workers." In: *Proceedings of the 23rd USENIX Security Symposium*, 239–254.

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H. and B. Y. Zhao (2012). "Serf and turf: crowdturfing for fun and profit." In: *Proceedings of the 21st International Conference on World Wide Web*, 679–688.

Westerman, D., Spence, P. R. and B. Van Der Heide (2012). "A social network as information: The effect of system generated reports of connectedness on credibility on Twitter." *Computers in Human Behavior* 28 (1), 199–206.