From Digital Footprints to Facts: Mining Marketing Policies of the Greek Community on Instagram and Youtube

Dimitris Linarakis, Stefanos Vlachos and Paraskevi Raftopoulou

Dept. of Informatics and Telecommunications, Univ. of the Peloponnese, Greece

Keywords: Social Media, Big Data, Data Analysis, Data Mining, Influencer Marketing, Covid-19, Greek Community, Instagram, YouTube, Web Scraping, NoSQL.

Abstract: Social media have conquered the widest spectrum of our lives. Companies, following the phenomena of the new digital era, are giving up traditional practices and use new policies of diffusion for advertising products and for engaging potential customers. In the context of ENIRISST+ (https://enirisst-plus.gr), we started investigating whether transportation businesses (i.e., ferry companies, airlines, etc.) use the new era practices for promoting their products and services and then, extended our research on all kind of businesses. The work presented in this paper studies this shift for the Greek Instagram and YouTube community, records and analyses the activity of prominent Greek companies on social media, and measures the social and commercial impact of the emerging COVID-19 pandemic during 2020 on Greek users’ digital behaviour. Subsequently, we use the acquired data and analysis (i) to draw conclusions about the digital behaviour and preferences of the Greek social media scene and (ii) to compare the results in marketing policies and behavioural patterns on two inherently different social media platforms. This is the first study in the literature to perform an analysis on the behaviour of the Greek community in different social media.

1 INTRODUCTION

The last twenty years, with the evolvement of digital services and the explosion of social networks, many aspects of our lives as well as our way of thinking and acting have been radically affected. Social networks hold a central role in interpersonal relationships, in the dissemination of information, knowledge and technology, in the formation of trends, and in the adoption of opinions; an ever-increasing number of users engages with and is affected by a multitude of posts on Instagram, YouTube, or Twitter (Eirinaki et al., 2018; Peng et al., 2018). This phenomenon, along with the fact that most content on online social networks is user-generated, has prompted many businesses, not only to have digital presence on social media, but also to give up the traditional advertising practices and use new policies, fitting the customers’ engagement with the social platforms, for promoting their products and services.

Among the various modern marketing techniques stands influencer marketing (IM). IM is a process of diffusion over social networks (Wang and Street, 2018), according to which a company, that wants to market a product or service cost- and time-effectively, identifies a small number of users (influencers) with high social network impact and whose social network presence tallies with their products. These users are becoming the initial adopters of the product recommending it to their friends (followers), expecting that their friends will be influenced to purchase the product and in turn, influence their own friends, causing an influence dissemination (through the social network) and eventually the wide adoption of the product. A recent survey (Advertising, 2019) shows that 80% of the consumers made purchases recommended by influencers. The work in (Vrontis et al., 2021) is a systematic review that identifies the key themes and the dominant concepts of social media IM. Moreover, (Chen et al., 2021) shows that platform characteristics significantly affect consumers’ purchase behaviour.

Instagram is the most used platform in the fashion industry. A large number of studies (Boerman and Müller, 2021; Casaló et al., 2020; Djafarova and Rushworth, 2017; Lin et al., 2019; Djafarova and Bowes, 2021; Asdecker et al., 2021) address influence dissemination and marketing under different perspectives attempting to gain insights on Instagram marketing tools, on the influencers’ personalities and on the consequences of IM on the purchase behaviour. The
last couple of years, the YouTube platform has also evolved as a primary marketing selection for brands promotion. Related works (Xiao et al., 2018; Acikgoz and Burnaz, 2021; Schwemmer and Ziewiecki, 2018; Prasad, 2018; Carreon, 2019; Boerman and van Reiijmersdal, 2020; Coates et al., 2020) study the role of YouTube as a social media platform and its impact on creating visibility on brands, as well as how cues influence credibility evaluations of posted information.

In this context, we started our research by investigating whether transportation businesses (i.e., ferry companies, airlines, etc.) use the new era practices for promoting their products and services; there was not a relative study available so far. We then generalised our research on all kind of businesses, focussing though on Greek market as it has recently entered social platforms. In this work, we study and analyse the interplay of commerce with popular social media platforms discerning three fundamental axes: (i) the common users’ digital behaviour, (ii) the influencers’ digital behaviour, and (iii) the interest of companies to draw marketing policies grounded on these behaviours. Based on their popularity during the year 2020, their different characteristics, and the continuously increasing commercial interest for them, we selected Instagram and YouTube platforms as the main subjects of our study.

Our research process is unfolded in two stages: (i) data collection and storage and (ii) data analysis. At the first stage, two web spiders were developed, using the Scrapy framework,1 one for each platform: the spiders received as input an initial hand picked list of users and collected data from Instagram and YouTube. Consequently, the collected data was stored in a MongoDB2 NoSQL database. Due to the fact that some fields of interest were not available for retrieving, we prepared and gave out to a wide range (in terms of characteristics) of volunteers a questionnaire in order to enrich the scraped data. Then, we analysed, based on the qualitative and quantitative properties of the data, the activity of the Greek market.

The contributions of our work are as follows: (i) we present a qualitative analysis of the activity of the Greek community on social platforms; to the best of our knowledge, this is the first study concerning the behaviour of the Greek community on social networks, (ii) we identify the marketing policies and behavioural patterns in Instagram and Youtube, (iii) we compare the results of these two completely different, in terms of content and interaction, social platforms, and (iv) we examine the social and commercial impact of COVID-19 on social media users’ activity.

This work is the first step to build a general framework to collect and analyse data from social platforms with minimum user intervention and by using filters per country, per topic, and/or per social network.

In the rest of the paper, we discuss the data collection and storage in Section 2, followed by the data analysis process in Section 3 and the results of our research in Section 4. Finally, Section 5 concludes the paper and gives future directions.

2 DATA COLLECTION & STORAGE

We searched for Greek non-private accounts, which had uploaded content within 2020 and maintained at least 1K followers (resp. 5K subscribers) on Instagram (resp. YouTube).3 These guidelines were set based on the fact that non-private accounts house publicly accessible data and also, have the potential to monetise their posts by promoting or selling a product or service (Dreghorn, 2020). We developed automated software that extracted the credentials of popular Greek content creators from credible web sources. For scraping the initial accounts sample we used stargage,4 a marketplace that connects brands with content creators from different media, and stats.videos,5 which includes YouTube statistics for popular video creators per region. Moreover, we collected the names of Greek businesses having digital presence in social media from socialbakers,6 a site that includes social media insights per industry or region. Eventually, we collected 2,766 Greek Instagram profiles (784 Male, 926 Female, 806 Business, and 249 Other) and 2,068 Greek YouTube channels (782 Male, 318 Female, 314 Business, and 654 Other), which will be later used as the feed sample of our scraping algorithms. The set of un categorised (Other) accounts consists of profiles that are not administered by individuals and includes impersonal accounts, i.e. profiles administrated by groups of people (e.g., tv channels, magazines, etc.).

[Instagram Dataset]. An Instagram (IG) account consists of 10 attributes, including Username, Number of followers, etc. Also, IG provides the capability of uploading a variety of content types, including 24-hours posts (i.e., IG stories) and posts that remain

1https://scrapy.org
2https://www.mongodb.com
3The thresholds do not hold for business accounts.
4https://stargage.com/app/global
5https://stats.video/top/most-subscribed/
youtube-channels/greece/of-all-time
6https://www.socialbakers.com/statistics/youtube/
channels/greece
From Digital Footprints to Facts: Mining Marketing Policies of the Greek Community on Instagram and Youtube

visible on the user’s feed until deleted. Despite the fact that numerous techniques were available for collecting the above mentioned IG fields, the complexity of using an authorised IG API led to the use of our software. An automated web scraping mechanism was developed, using the Scrapy framework and the Python programming language, which follows the given steps: (1) Access to the usernames of IG accounts that were initially collected. (2) Generate the URLs and the request for the source code of each account. (3) Extract the JSON structure from the HTML source code of each account. (4) Extract the personal information related to each scraped account from the JSON structure. (5) Access to the first 12 posts of each account and collect their information; IG groups the posts of each account in subsets of 12. (6) Request for the next 12 posts of each account and collect their information. (7) Repeat Step 6 until all posts of the account have been scraped. (8) Group the data related to each account in data sets and store this in the database.

[YouTube Dataset]. A YouTube (YT) channel consists of 8 attributes, including Username, Number of subscribers, etc. In contrast to IG, the YT platform provides the YT Data API, which supports the interaction of developers with a wide variety of information related to YT creators and content, enriching also the retrieved data with the adittional fields (e.g., Channel/Video ID, or Channel/Video keywords), allowing us to have access to hard-to-reach data. The communication with the YT Data API was established via an automated mechanism that was developed using the Scrapy framework. In order to retrieve data from the API, an appropriate HTTP request was needed, so that the requested data is specified. The successive steps are as follows: (1) Access to the channel IDs of YT channels that were initially collected. (2) Generate a URL that requests from the API the set of general information about each channel. (3) Collect the desirable fields (4) Generate a URL that requests from the API the first subset of videos for each channel; YT Data API groups the videos of each channel in subsets of 50, which are interconnected using pointers. (5) Generate a URL that requests from the API information related to the scraped videos. (6) Collect the desirable fields. (7) Repeat Steps 3-5 until all the videos of each channel have been collected. (8) Group the data related to each account in datasets and store it in the database.

The code of our scraping mechanisms is available on GitHub.\footnote{https://github.com/DimitrisLinarakis/\, https://github.com/stefanos-vlachos} [Questionnaire]. The privacy policies of the IG and YT platforms, in addition to the relatively limited amount of information available for collection, prevented our access to fields, which could enrich and confer completeness on the final conclusions of the present study. It was thus, decided to create and distribute a questionnaire, using Google Forms, in order to both provide a deeper insight of the preferences of Greek social media users, compared to the more impersonal metrics that had already been collected, and help in assessing the effectiveness of the diverse promotion techniques on social media, such as the use of hashtags, the cooperation with other content creators or the paid promotion of content.

Our questionnaire aimed at different Greek users of IG and YT, in terms of age and social characteristics. In particular, the questionnaire consisted of 4 introductory questions, 23 questions related to IG and 20 questions related to YT. The questionnaire was answered by 270 people above the age of 18, including 187 women (69.3%), 82 men (30.4%) and 1 non-binary person, whilst 128 of the participants (47.4%) were from the 18-24 age group, 53 (19.6%) from the 25-34 age group, 66 (24.4%) from the 35-44 age group, 20 (7.4%) from the 45-54 age group and 3 (1.1%) were above the age of 55.

3 ANALYSIS PROCESS

IM is an emerging field and the research questions we are challenged to answer are the following.

• Which are the underlying factors that drive users of networking platforms to interact, at an ever-increasing rate, with the content of the platform at a daily basis?
• Which were the preferences of the Greek social media community during 2020? Are these preferences age or gender related?
• Which were the thematic categories and types of content that obtained significant attention by the users during 2020? Which of these categories strived in 2020, in terms of content production?
• What is the role of businesses in social media and which are the potential relationships developed between the various categories of content creators?

The breakout of COVID-19 during 2020 brought radical changes to people’s daily lives, prompting us also to examine the exact effects of COVID-19 on the Greek social media community.

[Raw Analysis Measures]. We are interested in addressing the factors that affect the performance of an
uploaded post, and, thus, the rate of the exerted influence. Primarily, the popularity of an account can be inferred from the number of followers/subscribers that this account maintains. Accordingly, the number of “Like” reactions on IG/YT uploads can provide a first insight of content dynamics. On the contrary, the number of “Dislike” reactions on YT videos can signify the negative influence of content on the audience. The number of views on IG and YT videos can be also considered as a popularity indicator. All these form the basic axis on which the content creators’ influence is measured and collated with the IG/YT fields we can detect potential correlations that further signify the influence rate of a creator. Finally, the questionnaire serves the need for an ancillary research tool for data collection, as follows: (i) Verify the already collected data, (ii) Address the reasons why users follow influencers, (iii) Assess the influence rate of funded promotions, (iv) Examine the relationship between the popularity of an account and its influence rate, (v) Capture the community’s preference on IG stories over feed posts, and vice versa, and (vi) Collate all the collected data with social media users’ gender (Male, Female, Non-Binary) and age (18-24, 25-34, 35-44, 45-54, 55+).

**[Performance Metrics].** We trace more complex correlations and enhance the accuracy of the primary observations, in terms of the influence rate and the mapping of the social and commercial relationships developed on social media (Sehl and Tien, 2020), by using the following metrics: For the IG data: (i) Engagement Rate Post measures the performance of a post and is calculated as \((\text{Post Likes} / \text{Account Followers}) \times 100\), (ii) Engagement Rate Comments Post provides a deeper insight into the performance of a post and is calculated as \([\text{Post Likes} + \text{Post Comments}] / \text{Account Followers}] \times 100\), (iii) Engagement Rate Views measures the performance of a video and is calculated as \((\text{Video Likes}/\text{Video Views}) \times 100\), and (iv) Average Days Between Posts reflects the average upload frequency of a content creator and is calculated as \(\text{Sum of days between successive posts}/(\text{Total 2020 Posts - 1})\). Accordingly, for the YT data: (i) Video Dislikes Ratio reflects the negative impact of a video on the community and is calculated as \(\text{Video Dislikes}/(\text{Video Likes} + \text{Video Dislikes})\), (ii) Video Engagement calculates the engagement of the community with a video and is given as \((\text{Video Likes} + \text{Video Comments})/\text{Video Views}\), (iii) Views Subscribers Ratio measures the average percentage of the subscribers of a channel that watch the uploaded videos and is calculated as Video Views.

**[Outliers Management].** During the analysis stage of the data, there were accounts and posts that gathered an eminent amount of reactions, steering the measurements away from the actual picture. In order to prevent this phenomenon, the distribution of all the available measures was thoroughly examined. This process highlighted the outliers in the dataset, which were then excluded to ensure the reliability of the final conclusions.

### 4 DATA ANALYSIS RESULTS

In this section, we extensively present the results of our research.

#### 4.1 Instagram Results

Concerning IG, it has firmly established itself in the Greek community and, especially, in the younger users, with the majority of which being active on the platform for more than two years (point of reference is the year 2020). However, the ongoing entry of more older users in the platform proves its increasing rate of influence on the whole age spectrum of the community. The following findings hold.

- IG is a platform on which female users obtain significant attention, as they gathered the largest number of followers within 2020, with male users following second in the ranking.
- The higher influence rate marked by female/male accounts signifies that platform users are more strongly related to personal accounts (Fig. 1).
- Businesses hold a strict attitude regarding the number of accounts they follow; similar is the attitude of the female accounts (Fig. 2).
- The higher influence rate marked by female/male accounts signifies that platform users are more strongly related to personal accounts (Fig. 1).
- Businesses hold a strict attitude regarding the number of accounts they follow; similar is the attitude of the female accounts (Fig. 2).
- The higher influence rate marked by female/male accounts signifies that platform users are more strongly related to personal accounts (Fig. 1).
- Businesses hold a strict attitude regarding the number of accounts they follow; similar is the attitude of the female accounts (Fig. 2).
- Male users stayed less active, maintaining a relatively low content production rate.
- In summer of 2020 male and female content creators highly boosted their activity level.
Moreover, the upload of posts related to lifestyle, travelling and fashion/beauty during 2020 was significantly frequent, proving that IG provides a fertile ground for the development of such content. However, each category of content creators prospered in different thematic categories as follows. Female creators and the business accounts presented high content creation rate in the fashion/beauty field, male creators showed a preference towards travelling content, while the uncategorised accounts created posts in a wider spectrum of thematic categories, including travelling, arts and entertainment.

Taking also into consideration the answers to our survey, we concluded that the Greek community was particularly interested in lifestyle, arts, travelling, and entertainment during 2020. Nevertheless, male and female social media users expressed some special interests, with the first showing a preference to sports and technology, while the latter preferring fashion/beauty and cooking. Moreover, remarkable was the presence of contests (also known as Giveaways) in the Greek social media community. The organisation of numerous Giveaways by content creators with various interests, in combination with the huge engagement of the audience (regardless of the users’ age) with these posts, led us to the conclusion that Giveaways were the most popular type of IG posts for the year 2020 (Fig. 4).

4.2 YouTube Results

Concerning YT, it is quite different from IG. The sharing and playback of videos form the main pillar of this platform, while it is gradually adopting many modern social media features, such as the capability of uploading text posts and highlights. Considering that YT pre-existed IG, the Greek audience seems to have a high degree of familiarity with the character of the platform, as users of all ages have been using it for more than two years (point of reference is the year 2020). Nonetheless, despite the multiannual presence of elder users on the platform, it appears that YT is being better absorbed by younger age groups. The engagement of Greek users’ daily lives with YT can be also inferred from the frequent watch of videos in the platform. However, the more time required to interact with the YT content, compared to IG, has resulted in the partial shift of users to the weekly use of the platform, as opposed to the daily browsing through IG.

The following findings hold.

- Male creators and uncategorised channels have a stronger presence when compared with the number of channels administrated by female creators and businesses.
- Male creators and uncategorised channels have the highest number of subscriptions.
- Greek businesses marked the lowest engagement rate with their audience within 2020, as well as the highest rate of “Dislike” reactions on their videos.
- Channels administrated by male and female users had a better engagement with their subscribers, signifying that personalised channels establish a stronger relationship with the community (Fig. 5).
- Newly established channels also marked high engagement rates, implying that newer content creators cover the modern needs of the Greek YT community more effectively (Fig. 6).
- Male creators and uncategorised channels were the most active on the platform.
- The overall production rate was significantly increased in 2020.
- Creators’ productivity did not meet the need of the community to engage with the uploaded content during the summer months.
Entertainment, comedy, news, gaming, and music were the thematic categories that included the highest number of uploaded videos, with the first three showing a remarkable growth.

It is worth mentioning that each category of creators distinguished itself in the creation of videos related to specific thematic categories. In more detail, a large share of videos uploaded by male creators and uncategorised channels was focused on comedy, gaming and music, female creators uploaded content mainly related to lifestyle, fashion/beauty and cooking, while businesses exhibited a more diverse character, since their uploaded content was mainly concerning technology, fitness, gaming, and comedy.

Concerning the community’s engagement with the thematic categories within 2020, it applies that (i) gaming, comedy, entertainment, arts and educational content comprised poles of great interest, (ii) news and politics videos received negative response, holding the record for the highest average number of “Dislike” reactions, (iii) gaming, entertainment, and people & blogs formed the fundamental thematic categories of videos created by channels established in 2020, (iv) videos that include the active participation of viewers (i.e., QnAs, challenges, “We’re reacting to your stories”), as well as exhibition of products (i.e., reviews, haul, unboxing) obtained significant attention (Fig. 7), (v) videos that feature highlights from the creator’s daily life (Vlog) were remarkably popular (Fig. 7), (vi) contest videos (Giveaways) were found to be less prevalent on YT than on the IG platform, but they did achieve a quite high engagement rate (Fig. 7).

4.3 Business and Influencer Marketing

Since one of the fundamental axes of our research is the study of business presence on social media, we thoroughly examined their activity on the platforms. It appeared that IG provides a fertile ground for the development of commercial relationships, as it has managed to attract a multitude of businesses, which, in fact, follow a fairly intense activity. On the other hand, numerous businesses are also present on YT, although to a more limited extend when compared to IG. Businesses exhibit the following common features in both platforms.

- Transportation businesses (i.e., ferry companies, airlines, etc.) hold a relatively small portion (<5% for IG and <3,5% for YT) of the Greek companies having presence on social platforms.
- Businesses present a conservative and professional character on social media platforms, since they mostly adhere a consistent activity weekly schedule and follow sparingly other user accounts (Figs. 2 and 8).
- Businesses remarkably increase their activity during holidays, celebrations and anniversaries, such as Easter, summer, Christmas or Valentine’s Day, indicating that these are periods of increased commercial interest (Fig. 3).
- Businesses have not achieved yet to establish a committed community, as it was shown that Greek social media users maintained a poor engagement rate with content uploaded by businesses during 2020 (Figs. 1 and 5).

In order for businesses to promote their content on social media more effectively, they either advertise it via
regular posts/videos or paid promotions, or have their content promoted by popular content creators (IM). In more detail:

- The exposure of the advertised content to a group of users that are not followers or subscribers of the corresponding business leads to low engagement rates as well as, in some cases, to users’ feeling of annoyance.
- The Greek IG community appeared to be more tolerant to advertisements compared to YT.
- Regular advertising via posts/videos or paid promotions was observed to exert a higher influence rate on elder social media users.
- The high rates of audience’s engagement with content that included product placement within 2020 was the reason why numerous businesses turned to popular content creators for promotion (Figs. 4 and 7).
- IM was observed to have greater influence on younger social media users.

As already discussed, the differences between IG and YT platforms have favoured the development of different thematic categories and content types on each platform. This has shaped the business tactics on social media as follows: (i) businesses on IG expressed their preference towards female content creators, who mostly uploaded content related to lifestyle, (ii) businesses on IG were mostly interested in fashion and beauty, (iii) at YT male content creators were at the centre of attention of businesses, and (iv) the interests of businesses on YT were gaming, science/technology, fashion/beauty, and travelling. Finally, it is worth mentioning that the activity of businesses on IG was influential across a wide age range, in contrast to YT, where businesses corresponded to the needs of a younger audience.

4.4 Impact of COVID-19

Year 2020 was the one that brought several major changes to peoples’ daily lives, as it was the period during which the COVID-19 virus broke out. In the context of seeking for a more accurate view of this impact, the following observations were made.

- COVID-19 led to a bigger shift of users towards social media, with a large part of the Greek community using a plethora of hashtags related to the pandemic; few popular hashtags were actually accompanied by a feeling of nostalgia and reminiscence (e.g., #throwback, #memories).
- It was noticed that the first wave of the pandemic (March 2020-May 2020) had a greater impact on the public, with the second phase (October 2020-December 2020) being associated with a weakened but noticeable awareness among users (Figs. 9 and 10).

• YT had a clear benefit from this situation, with a sharp increase in the number of uploaded videos throughout 2020 (Fig. 11).
- The increased activity of users on social media, in combination with the suspension of several business operations, provided an opportunity for businesses to enhance, or even build from scratch, their digital presence on social media platforms.
- Certain thematic categories were also boosted during the pandemic: gaming, entertainment, comedy and news & politics where sectors whose productivity increased significantly.

5 CONCLUSIONS AND FUTURE WORK

In this work, we have presented a study concerning the interference of commerce with popular social media platforms discerning three fundamental axes: (i)
the common users’ digital behaviour, (ii) the influencers’ digital behaviour, and (iii) the interest of companies to draw marketing policies grounded on these behaviours. This is the first study concerning the activity of the Greek community on Instagram and YouTube platforms, the activity of Greek companies on social media, and the impact of Covid-19 on Greek consumers’ digital behaviour.

This work can be further extended by investigating more social network platforms. Moreover, we envision a general framework for collecting and analysing data from social platforms with minimum user intervention and by using filters per country, per topic, and/or per social network. Having such a framework, it would be of great interest to compare data coming from different communities and generalise our conclusions concerning the interplay of businesses with popular social media platforms.

ACKNOWLEDGEMENTS

This work was supported in part by project ENIRISTST+ under grant agreement No. MIS 5047041 from the General Secretary for ERDF & CF, under Operational Programme Competitiveness, Entrepreneurship and Innovation 2014-2020 (EPAnEK) of the Greek Ministry of Economy and Development (co-financed by Greece and the EU through the European Regional Development Fund).

REFERENCES


