

Article



Does SEO Matter for Startups? Identifying Insights from UGC Twitter Communities

José Ramón Saura ¹, Ana Reyes-Menendez ^{1,*} and Chris Van Nostrand ²

- ¹ Department of Business and Economics, Rey Juan Carlos University, Paseo Artilleros s/n, 28027 Madrid, Spain; joseramon.saura@urjc.es
- ² UC Berkeley Extension, University of California, Berkeley, CA 94720, USA; chrisvannostrand@berkeley.edu
- * Correspondence: ana.reyes@urjc.es

Received: 24 September 2020; Accepted: 21 October 2020; Published: 23 October 2020



Abstract: In the present study, we analyzed User Generated Content (UGC) to measure the importance of Search Engine Optimization (SEO) for startups. For this purpose, we used several clustering algorithms to identify user communities on Twitter. The dataset contained a total of 67,126 tweets. A three-step UGC analysis process was applied to the data. First, a Latent Dirichlet allocation (LDA) was developed to divide the UGC-sample into topics. Next, a sentiment analysis (SA) with machine-learning was applied to divide the sample of topics into negative, positive, and neutral feelings. Finally, a textual analysis (TA) process with data mining techniques was used to extract indicators related to the SEO technique optimization in startups. The results helped us identify UGC communities in Twitter about SEO for startups and the main optimization indicators according to the feelings expressed in tweets. Our results also demonstrated that Black Hack SEO is not the most relevant strategy of positioning of digital marketing for startups and that, although this strategy is used by the startups, it is predominantly negatively perceived by SEO UGC communities.

Keywords: SEO; startups; User Generated Content; social communities

1. Introduction

In recent years, the Internet has become a widely used and ever-growing data source [1]. A large proportion of data is generated daily by users who connect to social networks and who use their mobile applications to express opinions on products or services, or on any specific topic [2,3]. In today's increasingly connected world, users obtain information on the Internet through search engines, such as Google, Yahoo!, Bing, Baidu, among others [4,5].

A search engine is a website that allows to make inquiries on any subject [6]. In just a few thousandths of a second, search engines return multiple results from web pages containing the searched keywords. If a business is not indexed in the Search Engine Result Pages (SERPs), it does not appear in user search results. The strategy to position a company's web page in search engines results is known as Search Engine Optimization (SEO) [7–9].

Therefore, SEO is a technique that consists of the optimization of indicators within (on-page) and outside of the web page (off-page). At the same time, the importance of the SEO positioning strategy around the world increases due to the massive use of search engines, the development of new technologies, and innovation in the business sector [9]. This combination of factors has led to the emergence of startups, companies based on technology and innovation, in a connected ecosystem [10–12].

In a phase, startups need digital marketing and SEO positioning to increase their reputation on the Internet [13,14]. At this stage, startups have to increase the impact of their projects while looking for funding to invest in digital marketing [15,16].

This newly emerged digital ecosystem is rich in User Generated Content (UGC) [17], i.e., the content daily generated by Internet users through social networks and digital platforms. Relevant research has analyzed UGC for digital marketing, SEO, Search Engine Marketing (SEM), App Store Optimization (ASO), or even Social Media Marketing (SMM) [7,8,18–20].

The ecosystem of digital marketing in startups is characterized by the value of knowing the indicators, factors, techniques, or future topics that can affect the results in the most used search engines such as Google or Yahoo! [6] and that can increase the results and profitability of companies in the digital environment [9].

Therefore, we proposed to study two objectives, on the one hand, to understand and explain the main positioning indicators related to SEO strategies according to the UGC in Twitter, and on the other hand, to identify the main communities of users on Twitter that develop content on the main SEO indicators [21–23]. It is well known that, in the startup industry, the knowledge of these factors could generate added value to startups [15], since they can appropriately position themselves in their industries and generate a competitive advantage [16].

Therefore, there is a growing need to understand the topics that arouse interest in the SEO communities on Twitter, a major channel for the sharing of information in this industry [6,7,24]. To explore this issue, Aswani et al. [9] used data visualization techniques and content analysis of what they defined as SEM, considering also the SEO strategy within the acronym of SEM. In particular, they developed content analysis of tweets with hashtags #SEO, #SEM, and #DigitalMarketing. The authors also measured the polarity [25–28] and sentiment [29–32] of the collected tweets. Saura and Bennett [33] proposed the analysis of UGC communities using a three stages method. Therefore, following Aswani et al. [9] and Saura and Bennett [33], we propose the following research question (RQ): *What are the dominant discussion topics published in UGC SEO communities for startups in Twitter*? (RQ1).

Analysis of the content published in specific communities in social networks has aroused a considerably scholarly interest [22,27]. Users in these UGC communities' comment, express their opinions, criticize, or create content that brings improvements to the community [28]. In the startup industry, these practices are common and can be analyzed by extracting insights that help to improve the strategies of startups in their digital marketing actions by the specialization of professionals who share their experiences through UGC communities. Identifying the size, relevance, shape and number of participants, as well as their leaders' opinions, are essential for the identification of the communities that should be followed by startups to obtain information related to SEO [21,29]. Accordingly, we propose the following RQ: *Is it possible to identify the structure of the UGC community network that participates in SEO for startups discussions in Twitter?* (*RQ2*)

Likewise, other authors such as Reyes-Menendez et al. [14] and Saura and Bennett [33] proposed using the sentiment analysis methodology to measure whether topics identified in UGC can be considered negative, positive, or neutral with respect to the object of study. This approach is based on the analysis of the comments of the users of these communities around specific topics. Following these research studies, we propose the following RQ: *Is it possible to identify SEO positioning topics for startups in UGC communities on Twitter and identify corresponding feelings (positive, negative, or neutral)?* (RQ3).

In the present study, our key aim was to identify indicators that add value to startup companies. To this end, we analyze UGC using methods such as sentiment analysis and topic modeling, and other relevant approaches extending the Aswani et al. [9] approach based on data visualization. Regarding the research originality, the authors have not been able to find research studies that approximate the analysis of SEO positioning strategies in startups using data mining and topic-modeling techniques as well as communities of UGC analysis, so the results can be considered an original contribution to the literature.

Therefore, the main aim of the present study was to explore the discussions on SEO so that to acquire an in-depth understanding of the dynamics of this niche industry for startups. Consequently, the primary purpose of this study is discovery, not hypothesis testing and not trying to control variables, but to discover them. These variables could be used in future studies to analyse SEO strategies in

startups as well as constructs for quantitative methods adding important theoretical contribution to the literature [33]. The insights generated by the proposed method of analysis can be used to evaluate the SEO industry in startups and compare whether SEO is really important and worthwhile. On collecting 67,126 tweets with hangtags #SEO and #Startups, we submitted this database to several types of analysis including advanced algorithms, such as Hyperlink-Induced Topic Search (HITS), the PageRank (PA) algorithm distribution, and Eigenvector centrality distribution algorithm (ECD) to identify UGC communities as well as Latent Dirichlet allocation (LDA), Sentiment Analysis (SA), and textual analysis (TA) methods.

The remainder of this paper is structured as follows. Section 2 covers the literature review. Section 3 presents the methodology development and Section 4 the analysis of results. The findings are summarized and discussed in Section 5 (Discussion) and Section 6 (Conclusions).

2. Literature Review

In the last decade, the ecosystem of startups and the strategies that these companies use to promote their products has aroused a considerable research interest. For example, [34] conducted a study about the development of startups in an academic environment. The authors focused on the analysis of interdisciplinarity and performed a case study of an educational startup case called UniStartApp. Furthermore, [35] analyzed the contribution of digital marketing to markets, importance of communication in integrated marketing, and creation of value in digital channels and the UGC comments.

SEO is based on the optimization of indicators that measure the relevance of a web page on the Internet [21]. Considering the relevance of these indicators, the search engines index them in the first positions of its SERPs, the second page, and beyond [22].

Such indicators are many and varied and include, among others, the Page Authority (PA), which measures the quality of the content of a website page, and the Domain Authority (DA), which measures the relevance of the domain based on the number of visits and links pointing to it from other web pages. Combining these parameters (PA and DA), the founders of Google, Larry page, and Sergey Brin created a ranking called PageRank, which was improved by the Moz company [23] based on PA and DA indicators.

Other important indicators that optimize a website for SEO include the repetition of the main keyword in title, description, and URL tags content [24–26]. These three indicators are shown in the SERPs when a user performs a search. If the itinerancy of the keyword in the content of a page increases, this leads to an increase of the possibility of that page being indexed in search engines [18,27].

There are also other indicators, such as the sitemap, a file in XML language, which shows the structural design and the information architecture of a website so that search engines can correctly track it, or the correct installation of the robots.txt file, what tells search engines what content should not be indexed into the SERPs [28,29]. Other relevant indicators include social meta-tags, which offer search engines additional information about the content of the website: Facebook meta tags are called Facebook Open Graph Data; for Twitter meta data, social tags are called Twitter Cards [30].

As mentioned above, in order to obtain quality traffic and publicize their products and services at an early stage of their development, startups need to generate impact and visibility of their businesses for search engines [25,31].

In order to increase their impact and potential, as well as to receive new investment, startups should adopt the domain and optimization of SEO strategies [32]. Startup founders usually hire qualified specialists or develop SEO strategies on their own through social platforms or through other digital channels, such as email marketing or social ads. These platforms are rich in UGC content, that is as indicated, the content developed by users to express their opinions and to make comments about different industries that usually are organized around social communities in social networks. Therefore, social platforms could be used to obtain a holistic picture about customer satisfaction in a specific domain [9].

In addition, [36] studied startups in the early-stage phase, focusing on the main steps in the creation of digital companies. The authors analyzed the main tools and techniques of digital marketing for startups, demonstrating the importance of increasing the visibility of corresponding projects on the Internet.

Using the same approach, [37] investigated the creation of a brand from scratch and highlighted guidelines and factors to success in the digital environment. This digital field was found to be characterized by techniques such as SEO, SEM, or digital techniques. Likewise, [38] conducted a study of women entrepreneurs in startups, discovering their professional skills and indicating their perceptions of the opportunity to create new companies based on social media sources and trends.

Similarly, [39] analyzed the relationship between business orientation and performance of startups. The results of this study showed that relevant factors in this domain are technological orientation, type of digital communication strategies selected, and social capital of the company.

Focusing specifically on SEO strategies, [22] conducted an experiment to increase the revenue from website advertising owing to the correct use of social networks and SEO as traffic channels. The authors focused the study in "Linkbuilding" tactic. Furthermore, in a study on the use of these SEO strategies in e-commerce, which assessed the efficiency of these search engine strategies, [40] related the search engine marketing to the financing capacity of the companies.

In addition, Saura et al. [15] concluded that SEO positioning and qualitative indicators such as quality traffic and user experience (UX) can increase the existence of these strategies in startups. Likewise, [41] indicated that, for startups to succeed, and in order to increase the chances of attracting quality traffic in digital environments and thus raising the curiosity of the UGC, a successful product idea should be found. Accordingly, [11] studied the reason why a large number of startups fail in their initial stage. The authors proposed new organizational strategies to effectively use factors such as time and resources, mentioning that there should be a correct digital promotion of products through Internet search engines as well as in UGC communities and among influencers.

3. Methodology Development

The methodological process developed in this study is based on the analysis of SEO in relation to the research questions posed. The present study is exploratory, rather than hypothesis-testing. Therefore, our main aim was not to explore the impact of certain variables, but to discover them and illustrate how they can be analyzed in future research using the proposed methods [42].

As mentioned previously, social media (SM) have become a new area for the investigation of both structured databases and unstructured databases. In previous studies, SM analysis has been effectively used to study stock price fluctuations, disease prevention, event monitoring, election result predictions, disaster management, brand management, public relations, public opinion polling, improvement of tourism services, solutions for global warming, or studies of promotions such as #BlackFriday [39,43–51].

In the present study, we used Twitter-based UGC to investigate whether startups should invest into the SEO strategies to respond to the proposed research questions and objectives in the introduction section. During data collection, a total of 78,022 tweets with hashtags #SEO and #Startups were downloaded from the public Twitter API over a period of 3 months (April–July 2019) following for this collection process research studies such as [33]. Afterwards, the database was filtered and cleaned to improve the robustness of the data. The final data sample was composed by a total of 67,126 tweets.

3.1. Social Network Modeling Algorithms

In the next step, we applied an LDA model to identify most discussed topics/themes. Following the approach proposed by [33], the LDA model was created. LDA is a state-of-art mathematical model that can divide a sample into topics. Thereafter, these topics were visualized in nodes to understand their weight. It should be highlighted that the LDA algorithm is an approach used in exploratory methodologies based on data discovery that aims to identify topics in structured or unstructured

databases developed that was initially proposed by Pritchard and Stephens [52] as a machine learning technique and was later expanded in use and improved by Blei et al. [53].

Likewise, the data were also submitted to several clustering algorithms such as: Hyperlink Induced Topic Search (HITS), the PageRank algorithm, and the Eigenvector centrality distribution algorithm (See results). HITS algorithm is a link analysis algorithm that rates links between communities developed by Kleinberg [49]. This algorithm is used to the web link-structures to discover and rank the webpages relevant for a particular search, in this study, those links are extracted from Twitter users accounts used like links [54,55]. The PageRank algorithm is similar to HITS algorithm, but in this study, we used PageRank as a way of measuring the social network importance score of each user's profile in Twitter. The Page Rank score for a given node is based on the links made to that node from other nodes [56]. The links to a given node are called the backlinks/in-degrees for that node [57]. This algorithm was developed by Larry Page and Sergey Brin [58] at Stanford University in 1996, as part of a research project about a new kind of search engine. It is known as the first algorithm used by Google search engine to index websites. Finally, the eigenvector centrality algorithm is a measure of the influence of a node in a network. A high eigenvector score means that a node is connected to many nodes who themselves have high scores. This measure is used in graph theory and developed by Landau [49]. Additional details relative to the methods used can be found in Section 4.

3.2. Sentiment Analysis

In addition, we performed sentiment analysis of the tweets. To this end, we first used a total of 349 tweets to train the sentiment analysis algorithm that works with machine learning in Python to obtain Krippendorff's alpha value (KAV) [54,55].

KAV should obtain a result equal to or above 0.667, so that the results indicate that the algorithm has been trained a sufficient number of times, although Krippendorff's indicates that the minimum KAV should be adjusted according to the weight of the conclusions. In this sense, a KAV high could be \geq 0.800 while a KAV between the measures 0.667–0.800 could be used to tentatively define and argue conclusions [54].

Sentiment analysis algorithms are broadly used in research to identify sentiments relative to the research purposes. Research such as [8] and [33] have used sentiment analysis algorithms to extract insights from UGC databases on social networks such as Twitter, Facebook, or TripAdvisor.

This process allowed us to categorize the tweets into three groups (positive, negative, and neutral) depending on the sentiment expressed in them. This was followed by textual analysis of the data that helped us identify indicators related to SEO in startups.

3.3. Textual Analysis

In this step, we followed [33]. They stated that the data that make up the databases should be analyzed for the data mining process in order to extract insights. In this case, the process is based on the NVivo software, a qualitative analysis tool.

We selected to use the NVivo software as it does not require a specialized knowledge and has a simple interface that allows the researcher to correctly classify and structure the database in nodes. The data entry processes were manually performed in NVivo, although the databases were already divided into sentiments. In this case, a structure of nodes was created in which the words identified as connectors, prepositions, articles, and plural forms are filtered [33].

Subsequently, the nodes are defined as data containers grouped according to their characteristics. The structure and design of new nodes are used to group raw data as accurately as possible. An important indicator to perform a textual analysis process is the Weighted Percentage (WP) that shows the number of times a node repeats its content in the database. NVivo was used to calculate the WP.

Research such as X and X have used textual analysis algorithms to exploratory analysis of the results of databases structured by UGC.

3.4. Modularity Report

Furthermore, to identify relevant communities in the studied UGC, we used the algorithm of data visualization and classification previously proposed by [56]. For the resolution of the results, [57] proposed an algorithm to group the results in communities of neurons or nodes based on their modularity. Modularity is a measure of the structure of networks designed to measure the strength of the division of a network into clusters or communities [56].

Networks with a high modularity have dense connections between the nodes within modules, but sparse connections between nodes in different modules. Modularity is often used in optimization methods to detect the community structure in networks [46,48].

4. Results

Descriptive statistics provide an overview of the nature of the tweets, the users that interact through them, and the degree of engagement of relevant stakeholders [33]. In our dataset, of a total of 67,126 tweets, 64,145 were original tweets, 29% were replies to these tweets, and 27,459 were retweets (RT). These metrics highlight a very active interaction between the stakeholders related to the SEO community and startups. Likewise, a total of 15,176 different hashtags were detected in the sample, with a total of 15,494 unique users around this UGC.

Over 45% of the tweets contained more than one hashtag. Considering that a total of 15,494 unique users were identified, each user published on average 4.4 tweets, including 2.1 original tweets, 1.3 retweets, and 1 reply. Regarding the visibility of users, data analysis demonstrated that most users were active and visible in this social network, so the analyzed content can be assumed to yield relevant insights. In the tweets, there were a total of 51,290 different URLs.

A closer consideration of the data showed that the most popular words used in the discussions in the tweets (excluding SEO, startups, and digital marketing) were marketing (3624), business (3532), Google (3073), content (2679), search (2649), website (2253), digital (1892), and tech (1694). Similarly, the analysis of the hashtags demonstrated that a total of 766 unique hashtags were found in the tweets, and they appeared 15,176 times. The most repeated hashtags were #SEO (12.453), #startups (7391), #digitalmarketing (3097), #marketing (3050), #socialmedia (2385), and #business (1762) (See Figure 1).



Figure 1. Most popular words and hashtags found in the dataset.

A more in-depth analysis of the data demonstrated associations between hashtags, words, and users. First, popular terms included the words directly related to the best SEO practices for startups (e.g., trick, tutorial, tips, check, now, why, great, sharing). Second, frequently used were also the names of active companies and relevant tools in the sector (e.g., Google, Moz, Gmail, Pay per Click (PPC), Cost per mille (CPM), Business to Business (B2B), Business to Consumer (B2C), eCommerce, Screamingfrog, Google Search Console, Google Analytics, Google Ads, etc.). These findings highlight the strong interest of the UGC community to these tools and companies.

4.1. UGC Communities' Results

Figure 2a,b show the PageRank (PR) distribution of the identified communities [58]. The PR is an iterative algorithm that measures the importance of each node within the identified network.

This metric assigns each node a probability of being clicked many times. In addition, we also used the Hyperlink-Induced Topic Search (HITS), also known as Hubs Authority algorithm, which rates the authority and the hubs distribution [59] (Figure 2c,d).

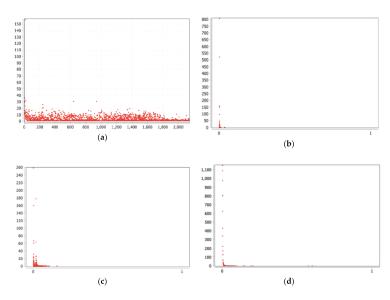


Figure 2. (**a**) Modularity distribution of communities; (**b**) PR and relevance nodes analysis; (**c**) Hubs distribution; (**d**) Hyperlink-Induced Topic Search (HITS) authority analysis (Source: The authors).

The HITS metric determines two values for a node: (1) its authority, which estimates the value of the content of the node, and (2) its hub value, which estimates the value of its links to other nodes. HITS updates the authority value of each node to be the sum of the hub values for every node it has a link to [59].

In our data, a total of 2145 user communities interacting on the topics related to SEO in startups and digital marketing strategies were identified after using the modularity report algorithm. In Figure 2a, the size (number of nodes) is shown on the Y axis, while the modularity class is shown on the X axis. The detected communities included communities of women in technology, volunteers and freelancers to develop SEO in startups and SEO, as well as SEM and SQL experts debating on the strategies and best advice for this industry. Of all these communities, the communities with more weight were as follows: SEO (the PR measure of 0.0326); business (0.0093); marketing (0.0086); startups (0.0070); digital marketing (0.0062); entrepreneurship (0.0039); innovation (0.0039); artificial intelligence (IA) (0.0037); social media marketing (SMM) (0.0036), and Fintech startups (0.0035).

Figure 2c shows hubs distribution where the points of nodes that stand out in the distribution correspond to Google (0.0594); ecommerce (0.0686); small business (0.0630); IoT (0.0606); SEM (0.043); B2B (0.0536); Blog (0.0575); PPC (0.0437); growth hacking (0.0607); analytics (0.0584); domains (0.0421); directory (0.0202); link building (0.0285); venture capital (0.0347); data (0.0487); founders (0.0305), and leadership (0.0487). These results demonstrate the interconnection of SEO tools and digital business models that develop startups.

Figure 2d shows the results on hubs authority. According to the results of this algorithm, the 10 nodes with the most authority within the sample are startups (0.5898); SEO (0.5739); business (0.2037); marketing (0.1917); entrepreneurship (0.0877); social media (0.0850); technology (0.0973); innovation (0.0834); design and web design (0.0605), and content marketing (0.0522). Two new communities focused on web design and innovation and content marketing strategy appear as well. This adds value to the results, suggesting that SEO can be applied to these digital areas. Y axis shows the Modularity weight of the communities, and the X axis represents the number of communities found by each algorithm used in Figure 2a–d.

4.2. Groups of UGC Communities

In order to better visualize the results of the processes depicted in Figure 2a–d, Figure 3 shows the UGC communities in terms of the weight of the corresponding nodes. As can be seen in Figure 3, of a total 2145 identified communities, 18 had a greater weight. By order of weight, our results showed that the node/community corresponding to SEO, business and marketing had the weight of 0.05062; startups, digital marketing, and entrepreneurship had the weight of 0.0454. These communities were followed, in the descending order of weight, by innovation and AI (0.0077); Fintech startups and blockchain startups (0.0059); SMM and content marketing (0.0058). Communities from C1 to C14 contained indicators of SEO optimization, tools, and tips related to SEO; these were small user communities focused on the discussion of the best practices and offering SEO advice for startups.

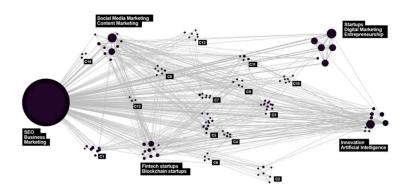


Figure 3. Groups of communities associated by similar content. Source: The authors.

4.3. LDA Results

Next, the main aim of the process of identifying topics with LDA was to find SEO optimization indicators for startups. The identified topics are listed in Table 1.

Торіс	Brief Description	Count
Robots.txt and Tag	Robots.txt and tag should be implemented on the web indicating the URL to the sitemap.xml; in its tag, it should appear as "index, follow".	161
URL	URLs should contain the keyword that positions the website; no noise or special characters are allowed.	157
Title Tag	Title tag should not exceed 69 characters and must contain the keyword.	139
Social Tags	Social tags should independently appear on all pages of the website.	123
Open Graph Data	Open Graph Data are used to mark Facebook data for social tags. It is important to optimize the "og: image".	110
JavaScript (JS)	There are problems in the implementation of JS for SEO	104
Backlinks	Backlinks are necessary to increase the PA and DA of the website, but there are penalties for their purchase.	104
Twitter Card Data	Twitter Card Data serve to optimize social SEO for Twitter that increases the impact of content in this social network.	98
Sitemap.xml	Sitemap.xml structures the content of the website and offers search engines the information architecture of the site.	92
Meta Description	Meta description should contain the keyword to be positioned and should not exceed 156 characters for the results in desktop SERPs and 333 for mobile.	83
Traffic	Traffic of a website relevant to SEO should be high.	79
Accelerated Mobile Pages (AMP)	Accelerated Mobile Pages (AMP) is a technology that increases the loading time of a website optimized to SEO.	79
Long-tail Keywords	Long-tail keywords are the most successful keywords in SEO	71
SEM	SEM is a digital marketing strategy that should accompany SEO tactics.	67
Directory	A website structure should be divided into directories and not in URLs by subject or sub-domain.	63
Link Building	Link building is a strategy that penalizes SEO.	57
CTR	CTR is the indicator that measures the effectiveness of clicks on SERPs.	35
Headers (H1-H6)	Headers show the search engines the structure of the content.	15
HTML	The goals tags in HTML and social tags are important for optimization	14

Table 1. Latent Dirichlet allocation (LDA) topics identification of Search Engine Optimization (SEO)
indicators for startups.

On identification of the topics, we submitted the data to sentiment analysis to identify the predominant feelings expressed in the tweets (positive, negative, and neutral). To this end, we first trained a machine learning algorithm developed in Python in three feelings (positive, negative, and neutral). The algorithm was trained on a total of 349 samples, and the average KAV values achieved for positive, negative, and neutral sentiments were 0.721, 0.775, and 0.801, respectively. Considering the conventional thresholds for KAVs values ($\alpha \ge 0.800$ high reliability; $\alpha \ge 0.667$ tentative conclusions; $\alpha < 0.667$ low reliability; see Krippendorff [54]), highly reliable and tentative conclusions could be made based on our data.

To better visualize the findings, Figure 4a,b show two word-clouds with the main words arranged according to their weight and feeling (positive and negative; neutral was not relevant in this case). It is important to highlight that Figure 4 shows the keywords found as a result of the process randomly.

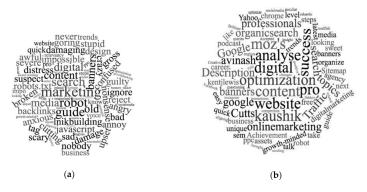


Figure 4. (a) Negative topics for SEO startups; (b) positive SEO topics for startups. Source: The authors.

As can be observed in Figure 4b, positive issues that should optimize the startups are title, description, URL, AMP, sitemap.xml, long-tail keywords, traffic, and social tags. Furthermore, as shown in Figure 4a, negative issues are robots.txt and tag, JavaScript, backlinks, and link building. There have also been figures of influencers. In addition, influential figures in the SEO sector, such as Rand Fishkin, Avinash Kaushik, Matt Cutts, or Eduardo Garolera, turned out to be positively evaluated in our sample.

5. Discussion

Previous studies have convincingly demonstrated that SEO is beneficial for startups that do not use SEM in their strategies to increase their visibility in the top of the SERPs [31]. Previous research has also identified the key factors that affect the positioning of startups in SERPs and central concepts related to user behavior or psychology [28]. In addition, SEO has been demonstrated to be an effective strategy to improve traffic to websites, which also brings profit [29]. Likewise, several studies have also investigated the impact of SEO strategies on global digital marketing campaign and pointed out that search marketing strategies are not as profitable as other strategies carried out by advertisers in the digital ecosystem.

However, none of previous studies have focused on the importance that startups should assign to SEO positioning, particularly in relation to the optimization of their social media marketing, SEO, and content marketing strategies. To fill this gap in the literature, in the present study, we aimed to explore the startup industry through a review of Twitter discussions.

Our results demonstrate that many small startup communities have conversations about digital marketing and the main strategies that should be implemented as demonstrated by [29]. In addition, our findings also clearly indicate that startups are closely interconnected through the most important user and thematic communities where they find useful information about SEO optimization [60].

Therefore, as indicated by Saura et al. [8] and Reyes-Menendez et al. [14], the data published on Twitter and the study of online communities can be a valid source of information to understand

public opinion, to find insights or even to predict social movements. In this way, the analysis and representation of clusters from Twitter can contribute to the identification of success indicators relative to different industries as indicated by Ventocilla [61]. The study of online communities and the topics on which users share information could also help to identify the level of engagement and opinion leaders, among other social listening indicators [62].

Therefore, based on our results about the network dynamics and clusters, it can be concluded that the industry concentration appears to be high, although highly fragmented (see Aswani et al. [9] for a similar conclusion). One of the reasons behind this trend may be the close contact among startups, as they are in the same phase of development and have on the same needs [25]. In addition, Fishkin [6] indicates that certain success factors for indexing content in the Google search engine are born from professional comments from users on Twitter when recounting their experiences

However, for the success of the SEO strategy in startups, the identified indicators should be optimized. The presence of negative SEO indicators suggests that there are negative experiences and bad results of other startups and practitioners in the industry as presented by [27]. Several previous studies have analyzed the negative engagement in digital marketing based on the most practical type of black hat SEO (see [63] and Aswani et al. [9]).

In addition, our results show that SEO is not always the strategy that should be carried out in terms of better positioning of startups in search engines (see also Malaga [27] and Aswani et al. [9] for similar conclusions).

Instead, alternative strategies, such as PPC, SEM, content marketing, SMM, or influencers marketing, can be meaningfully used to develop the positioning focused on startups [1]. However, despite the short-term gains in terms of attracting traffic, in the long run, the links can become toxic and the purpose behind the whole exercise can be defeated, resulting in no current gain of online. The aforementioned issues should become the top concerns for the startup industry [64].

6. Conclusions

In the present study, we focused on investigating startup SEO strategies of digital marketing. Our results on the discussions surrounding SEO on Twitter highlighted that 27% of all discussions had a negative polarity, indicating that SEO is not the perfect and profitable strategy for startups. This finding also suggests that there is a high percentage of unsatisfied SEO experts' experiences and negative engagement. The analysis clearly demonstrated that most users are not satisfied with the performance of their SEO strategies in startups.

A detailed analysis revealed that the major reason behind the dissatisfaction was outsourcing of negative SEO indicators and topics, such as robots.txt and tag, JavaScript, back links, and link building. In the long run, using such techniques or wrong optimization, when detected by a search engine, leads to penalization.

Along with contributing to the previous literature [65,66] our results provide meaningful practical implications for startups, organizations, and individuals searching for the ways towards quick and efficient solutions to enhance their web visibility. In addition, regarding research objective one, the main positioning indicators related to SEO strategies according to the UGC in Twitter have been explained in Table 1 allowing be clearly understood by practitioners or academics. Likewise, regarding research objective 2, the communities of users on Twitter that develop content on the main SEO indicators have been identified and scored. This processed let us find main topics for future research in this area.

However, studies using UGC to identify insights for startups to appropriately develop their digital marketing strategies remain scarce. The present study fills this gap in the literature. As concerns our first research question, RQ1 (*What are the dominant discussion topics published in UGC SEO communities for startups in Twitter*?), this study has shown that the main topics established by UGC in Twitter are these: SEO, business and marketing, startups, digital marketing entrepreneurship, innovation, AI, Fintech startups, blockchain startups, SMM, and content marketing. At the same time, the main topics are SEO, business and marketing, startups, digital marketing, and entrepreneurship.

However, important insights can also be obtained regarding the analysis of the communities linked to innovation, such as the development path of SEO strategies; AI, a technology that can be applied in the medium-long term to SEO strategies; Fintech startups, a business niche where startups can develop their SEO strategies; blockchain startups, specifically for the application of strategies focused on exploiting this technology; and SMM and content marketing, two key techniques for the success of startups.

Regarding RQ2 (Is it possible to identify the structure of the UGC community network that participates in SEO for startups discussions in Twitter?), it has been confirmed as the communities visually represented by their weight and size, as can be identified as seen in Figure 3. The weight of these communities has been measured using the Modularity Report indicator. As a result, startups can understand the relevance of each community for SEO positioning and target different communities.

Finally, regarding RQ3 (Is it possible to identify SEO positioning topics for startups in UGC communities on Twitter and identify corresponding feelings (positive, negative, or neutral)?), as indicated in RQ1, topics have been detected as a result of the textual analysis approach. The identification of sub-topics that determine the SEO optimization indicators have been as well divided into sentiments. Positive indicators are title, description, URL, AMP, sitemap.xml, long-tail keywords, traffic, and social tags. Negative indicators found are robots.txt and tag, JavaScript, backlinks, and link building. Additionally, among the positive indicators, prestigious figures in the SEO industry were identified such as Rand Fishkin, Avinash Kaushik, Matt Cutts, or Eduardo Garolera.

6.1. Implications for Practitioners

The results of the present study offer several insights for practitioners in the startup industry. First, startups can use our results to improve the SEO topics and indicators identified in the Twitter-based UGC. In addition, they can also obtain information regarding the communities that surround the digital marketing environment on Twitter and therefore use them to get insights that could help to develop their SEO strategies.

Startups can also use our findings to measure their impact in terms of performance and identify influencers that can help them improve their impact through the engagement in social networks. Furthermore, our study provides a global and dynamic vision of the startup ecosystem for industry practitioners. Finally, the insight suggested by our findings is that digital agencies and freelancers can improve their startup performance by being active in the identified communities and trying to increase their impact by using white hat SEO strategies. In future research, the methodology used in the present study can be expanded by professional studies to analyze other communities in the startups sector.

6.2. Theoretical Implications

As a result of the two processes outlined here, an important theoretical implication is the study of the 19 topics directly linked to the SEO environment in Startups industry.

If researchers take these insights as variables and constructs for their quantitative models, they may be able to enhance their understanding of whether positive links exist between them by developing, for example, models based on Partial Least Squares Structural Equation Modeling (PLS-SEM) or SPSS, Analysis of Moment Structures (AMOS), among others, thus contributing to a field of research that emerges from approaches that extract knowledge from large amounts of data. For example, does Social Tags identified topics influence web site traffic? Does CTR influence positively Linkbuilding strategies in startups?

In addition, academics can use this research to better understand the startups sector and to focus on the development of research within the field. In addition, they can focus on content analysis by users of different social networks to better understand what the main habits are when sharing information publicly in digital ecosystems.

6.3. Limitations and Future Research Directions

This study is based on Twitter to measure the importance of SEO strategies in startups. To complement our findings, future studies should consider including other social platforms in their analyses. In addition, in future research, it would make sense to broaden the time horizon. The methodological approach based on content analysis, analysis with algorithms, social analytics, topic-modeling, and sentiment analysis used in the present study can also be extended to obtain additional findings.

Author Contributions: Conceptualization J.R.S., A.R.-M. and C.V.N.; methodology, J.R.S.; software, J.R.S.; validation; A.R.-M. and C.V.N.; formal analysis, A.R.-M. and C.V.N.; investigation J.R.S., A.R.-M. and C.V.N.; writing—original draft preparation, J.R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Yoo, C.; Yang, D.; Kim, H.; Heo, E. Key Value Drivers of Startup Companies in the New Media Industry-The Case of Online Games in Korea. *J. Media Econ.* **2012**, *25*, 244–260. [CrossRef]
- Regmi, K.; Ahmed, S.A.; Quinn, M. Data Driven Analysis of Startup Accelerators. *Univers. J. Ind. Bus. Manag.* 2015, 3, 54–57. [CrossRef]
- 3. Rathore, A.K.; Ilavarasan, P.V.; Dwivedi, Y.K. Social Media Content and Product Co-Creation: An Emerging Paradigm. *J. Enterp. Inf. Manag.* **2016**, *29*, 7–18. [CrossRef]
- Paik, Y.; Woo, H. Economic Downturn and Financing Innovative Startup Companies. *Manag. Decis. Econ.* 2014, 35, 114–128. [CrossRef]
- 5. Thiel, P.A.; Masters, B.G. Zero to One: Notes on Startups, or How to Build the Future.; Virgin Books: London, UK, 2015.
- 6. Fishkin, R. SEO: The Beginner's Guide to Search Engine Optimization from Moz. Moz 2015. Available online: https://bit.ly/2UKj7U5 (accessed on 10 February 2019).
- 7. Reyes-Menendez, A.; Saura, J.R.; Filipe, F. The importance of behavioral data to identify online fake reviews for tourism businesses: A systematic review. *Peerj Comput. Sci.* **2019**, *5*, e219. [CrossRef]
- 8. Saura, J.R.; Reyes-Menendez, A.; Thomas, S. Gaining a deeper understanding of nutrition using social networks and user-generated content. *Internet Interv.* **2020**, 100312. [CrossRef]
- Aswani, R.; Kar, A.K.; Ilavarasan, P.V.; Dwivedi, Y.K. Search engine marketing is not all gold: Insights from Twitter and SEOC. In *Software Development in Startup Companies: The Greenfield Startup Model. IEEE Transactions on Software Engineering*; Giardino, C., Paternoster, N., Unterkalmsteiner, M., Gorschek, T., Abrahamsson, P., Eds.; IEEE: Piscataway, NJ, USA, 2016; Volume 42, pp. 585–604. [CrossRef]
- 10. Reyes-Menendez, A.; Saura, J.R.; Stephen, B.T. Exploring Key Indicators of Social Identity in the #MeToo Era: Using Discourse Analysis in UGC. *Int. J. Inf. Manag.* **2020**, *54*, 102129. [CrossRef]
- 11. Saura, J.R. Using Data Sciences in Digital Marketing: Framework, Methods, and Performance Metrics. *J. Innov. Knowl.* **2020**, *1*. [CrossRef]
- 12. Kropp, F.; Lindsay, N.J.; Shoham, A. Entrepreneurial Orientation and International Entrepreneurial Business Venture Startup. *Int. J. Entrep. Behav. Res.* **2008**, *14*, 102–117. [CrossRef]
- Davila, A.; Foster, G.; Li, M. Reasons for Management Control Systems Adoption: Insights from Product Development Systems Choice by Early-Stage Entrepreneurial Companies. *Acc. Organ. Soc.* 2009, 34, 322–347. [CrossRef]
- 14. Aswani, R.; Kar, A.K.; Ilavarasan, P.V.; Dwivedi, Y.K. Search engine marketing is not all gold: Insights from Twitter and SEOClerks. *Int. J. Inf. Manag.* **2018**, *38*, 107–116. [CrossRef]
- 15. Dellermann, D.; Ebel, P.; Lipusch, N.; Popp, K.M.; Leimeister, J.M. Finding the Unicorn: Predicting Early Stage Startup Success through a Hybrid Intelligence Method. In *ICIS 2017: Transforming Society with Digital Innovation*; Association for Information Systems: Atlanta, GA, USA, 2018. [CrossRef]
- 16. Giardino, C.; Wang, X.; Abrahamsson, P. *Why Early-Stage Software Startups Fail: A Behavioral Framework;* Springer: Cham, Switzerland, 2014; pp. 27–41. [CrossRef]

- Giardino, C.; Bajwa, S.S.; Wang, X.; Abrahamsson, P. Key Challenges in Early-Stage Software Startups. In *Lecture Notes in Business Information Processing*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 212, pp. 52–63. [CrossRef]
- Davila, A.; Foster, G. Management Control Systems in Early-Stage Startup Companies. Acc. Rev. 2007, 82, 907–937. [CrossRef]
- Garcia-Perez, A.; Romero-Troncoso, R.J.; Cabal-Yepez, E.; Osornio-Rios, R.A.; Rangel-Magdaleno, J.D.J.; Miranda, H. Startup Current Analysis of Incipient Broken Rotor Bar in Induction Motors Using High-Resolution Spectral Analysis. In Proceedings of the SDEMPED 2011—8th IEEE Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives, Bologna, Italy, 5–8 September 2011; pp. 657–663. [CrossRef]
- 20. Nanda, R.; Rhodes-Kropf, M. Investment Cycles and Startup Innovation. J. Financ. Econ. 2013, 110, 403–418. [CrossRef]
- Hettiarachchi, S.; Spears, W.M.; Varghese, B.; Mckee, G.; Evans, M.P.; Tokhi, O.; Bouloubasis, A.; Tolson, P. Analysing Google Rankings through Search Engine Optimization Data. *Int. J. Intell. Comput. Cybern.* 2007, 17, 116–121. [CrossRef]
- 22. Zhang, S.; Cabage, N. Search Engine Optimization: Comparison of Link Building and Social Sharing. *J. Comput. Inf. Syst.* **2017**, *57*, 148–159. [CrossRef]
- 23. Krrabaj, S.; Baxhaku, F.; Sadrijaj, D. Investigating Search Engine Optimization Techniques for Effective Ranking: A Case Study of an Educational Site. In Proceedings of the 2017 6th Mediterranean Conference on Embedded Computing, MECO 2017—Including ECYPS 2017, Bar, Montenegro, 11–15 June 2017; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2017. [CrossRef]
- 24. Lui, R.; Au, C.H.I.S. Educational Game: Adoption in Teaching Search Engine Optimization (SEO). J. Comput. Inf. Syst. 2020, 60, 233–243. [CrossRef]
- 25. Lui, R.W.C.; Au, C.H. Establishing an Educational Game Development Model: From the Experience of Teaching Search Engine Optimization. *Int. J. Game-Based Learn.* **2018**, *8*, 52–73. [CrossRef]
- 26. Pidpruzhnikov, V.; Ilchenko, M. Search Optimization and Localization of the Website of Department of Applied Linguistics. In Proceedings of the Computational linguistics and intelligent systems (COLINS 2017), Lviv, Ukraine, 21 April 2017.
- 27. Malaga, R.A. *Search Engine Optimization—Black and White Hat Approaches;* Elsevier: Amsterdam, The Netherlands, 2010; Volume 78, pp. 1–39. [CrossRef]
- 28. Cahill, K.; Chalut, R. Optimal Results: What Libraries Need to Know about Google and Search Engine Optimization. *Ref. Libr.* 2009, *50*, 234–247. [CrossRef]
- 29. Giomelakis, D.; Veglis, A. Investigating Search Engine Optimization Factors in Media Websites: The Case of Greece. *Digit. Journal.* **2016**, *4*, 379–400. [CrossRef]
- Aswani, R.; Kar, A.K.; Aggarwal, S.; Vigneswara Ilavarsan, P. Exploring Content Virality in Facebook: A Semantic Based Approach. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 10595 LNCSM; Springer: Berlin/Heidelberg, Germany, 2017; pp. 209–220. [CrossRef]
- 31. Baye, M.R.; de los Santos, B.; Wildenbeest, M.R. Search Engine Optimization: What Drives Organic Traffic to Retail Sites? *J. Econ. Manag. Strategy* **2016**, *25*, 6–31. [CrossRef]
- Farooqi, S.; Jourjon, G.; Ikram, M.; Kaafar, M.A.; de Cristofaro, E.; Shafiq, Z.; Friedman, A.; Zaffar, F. Characterizing Key Stakeholders in an Online Black-Hat Marketplace. In Proceedings of the 2017 APWG Symposium on Electronic Crime Research (eCrime), Phoenix, AZ, USA, 25–27 April 2017; pp. 17–27. [CrossRef]
- 33. Saura, J.R.; Bennett, D.R. A Three-Stage Method for Data Text Mining: Using UGC in Business Intelligence Analysis. *Symmetry* **2019**, *11*, 519. [CrossRef]
- 34. Palos-Sanchez, P.; Saura, J.R.; Martin-Velicia, F. A study of the effects of Programmatic Advertising on users' Concerns about Privacy overtime. *J. Bus. Res.* **2019**, *96*, 61–72. [CrossRef]
- 35. Singh, O.; Singh, K.A. Formulation of Value Proposition for Digital Marketing Strategy in Startups. *Natl. J. Multidiscip. Res. Dev.* **2017**, *2*, 45–52.
- 36. Lee, S.H.; Noh, S.H. A Study on the Success Factors of ICT Convergence Type-Specific Start-up Enterprise—Mainly the Case Study. J. Digit. Converg. 2014, 12, 203–215. [CrossRef]

- Lagerstedt, M.; Mademlis, A. Branding for Startup Companies in Sweden a Study on Startups Brand Building. Master's Thesis, University of Amsterdam, Amsterdam, The Netherlands, 2016.
- 38. Neill, S.; Metcalf, L.; York, J.L. Seeing What Others Miss: A Study of Women Entrepreneurs in High-Growth Startups. *Entrep. Res. J.* 2015, *5*, 293–322. [CrossRef]
- 39. Lee Hughes, A.; Palen, L. Twitter Adoption and Use in Mass Convergence and Emergency Events. *Int. J. Emerg. Manag.* **2009**, *6*, 248–260. [CrossRef]
- 40. Yang, Z.; Shi, Y.; Wang, B. Search Engine Marketing, Financing Ability and Firm Performance in E-Commerce. *Procedia Comput. Sci.* **2015**, *55*, 1106–1112. [CrossRef]
- 41. Bosch, J.; Olsson, H.H.; Björk, J.; Ljungblad, J. The Early Stage Software Startup Development Model: A Framework for Operationalizing Lean Principles in Software Startups. In *Lecture Notes in Business Information Processing*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 167, pp. 1–15. [CrossRef]
- 42. Corbin, J.; Strauss, A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory; Sage: Los Angeles, CA, USA, 2015.
- 43. Reyes-Menendez, A.; Saura, J.R.; Martinez-Navalon, J.G. The impact of e-WOM on Hotels Management Reputation: Exploring TripAdvisor Review Credibility with the ELM model. *IEEE Access* 2019, 8. [CrossRef]
- 44. Saura, J.R.; Herraez, B.R.; Reyes-Menendez, A. Comparing a Traditional Approach for Financial Brand Communication Analysis with a Big Data Analytics Technique. *IEEE Access* **2019**, *7*, 37100–37108. [CrossRef]
- 45. Arias, M.; Arratia, A.; Xuriguera, R. Forecasting with Twitter Data. *ACM Trans. Intell. Syst. Technol.* **2013**, *5*, 1–24. [CrossRef]
- 46. Chae, B. Insights from Hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter Data for Supply Chain Practice and Research. *Int. J. Prod. Econ.* **2015**, *165*, 247–259. [CrossRef]
- 47. Inauen, S.; Schoeneborn, D. Twitter and Its Usage for Dialogic Stakeholder Communication by MNCs and NGOs. Critical Studies on Corporate Responsibility. *Gov. Sustain.* **2014**, *6*, 283–310. [CrossRef]
- 48. Joseph, N.; Kar, A.K.; Ilavarasan, P.V.; Ganesh, S. Review of Discussions on Internet of Things (IoT): Insights from Twitter Analytics. *J. Glob. Inf. Manag.* 2017, *25*, 38–51. [CrossRef]
- 49. Landau, E. "Zur relativen Wertbemessung der Turnierresultate". *Dtsch. Wochenschach* **1985**, *11*, 366–369. [CrossRef]
- 50. Williams, S.A.; Terras, M.M.; Warwick, C. What Do People Study When They Study Twitter? Classifying Twitter Related Academic Papers. *J. Doc.* **2013**, *69*, 384–410. [CrossRef]
- 51. Wu, B.; Shen, H. Analyzing and Predicting News Popularity on Twitter. *Int. J. Inf. Manag.* 2015, 35, 702–711. [CrossRef]
- 52. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
- 53. Blei, D.M.; Ng, A.Y.; Jordan, M.I.; Lafferty, J. Latent Dirichlet Allocation. J. Mach. Learn. Res. 2003, 3, 993–1022.
- 54. Krippendorff, K. Agreement and Information in the Reliability of Coding. *Commun. Methods Meas.* 2011, 5, 93–112. [CrossRef]
- 55. Krippendorff, K. Bivariate Agreement Coefficients for Reliability of Data. *Sociol. Methodol.* **1970**, *2*, 139. [CrossRef]
- 56. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast Unfolding of Communities in Large Networks. *J. Stat. Mech. Theory Exp.* **2008**, *10*, P10008. [CrossRef]
- 57. Lambiotte, R.; Delvenne, J.-C.; Barahona, M. Laplacian Dynamics and Multiscale Modular Structure in Networks. *IEEE Trans. Netw. Sci. Eng.* 2008, 1, 76–90. [CrossRef]
- 58. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Stanford InfoLab: Stanford, CA, USA, 1999.
- 59. Kleinberg, J.M. Authoritative Sources in a Hyperlinked Environment. J. ACM 1999, 46, 604–632. [CrossRef]
- 60. Giomelakis, D.; Karypidou, C.; Veglis, A. SEO inside Newsrooms: Reports from the Field. *Future Internet* **2019**, *11*, 261. [CrossRef]
- 61. Ventocilla, E.; Riveiro, M. A comparative user study of visualization techniques for cluster analysis of multidimensional data sets. *Inf. Vis.* **2020**, *19*, 318–338. [CrossRef]
- 62. Kotras, B. Opinions that matter: The hybridization of opinion and reputation measurement in social media listening software. *Media Cult. Soc.* **2020**, *42*, 1495–1511. [CrossRef]
- 63. Ward, A.A. The Seo Battlefield: Winning Strategies for Search Marketing Programs; OReilly: Beijing, China, 2017.

- 64. Nguyen-Duc, A.; Shah, S.M.A.; Ambrahamsson, P. Towards an Early Stage Software Startups Evolution Model. In Proceedings of the 42nd Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2016), Limassol, Cyprus, 1 August–2 September 2016; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2016; pp. 120–127. [CrossRef]
- 65. Karyotakis, M.-A.; Lamprou, E.; Kiourexidou, M.; Antonopoulos, N. SEO Practices: A Study about the Way News Websites Allow the Users to Comment on Their News Articles. *Future Internet* **2019**, *11*, 188. [CrossRef]
- 66. Ziakis, C.; Vlachopoulou, M.; Kyrkoudis, T.; Karagkiozidou, M. Important Factors for Improving Google Search Rank. *Future Internet* 2019, *11*, 32. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).