Framework for Social Media Analysis Based on Hashtag Research

Ladislav Pilař [1],*, Lucie Kvasničková Stanislavská [1], Roman Kvasnička [2], Petr Bouda [1] and Jana Pitrová [1]

[1] Department of Management, Faculty of Economics and Management, Czech University of Life Sciences Prague, 165 21 Prague, Czech Republic; kvasnickova@pef.czu.cz (L.K.S.); boudap@pef.czu.cz (P.B.); pitrovaj@pef.czu.cz (J.P.)

[2] Department of Systems Engineering, Faculty of Economics and Management, Czech University of Life Sciences Prague, 165 21 Prague, Czech Republic; kvasnicka@pef.czu.cz

*   Correspondence: pilarl@pef.czu.cz

Abstract: Social networks have become a common part of many people's daily lives. Users spend more and more time on these platforms and create an active and passive digital footprint through their interaction with other subjects. These data have high research potential in many fields, because understanding people's communication on social media is essential to understanding their attitudes, experiences and behaviours. Social media analysis is a relatively new subject. There is still a need to develop methods and tools for researchers to help solve typical problems associated with this area. A researcher will be able to focus on the subject of research entirely. This article describes the Social Media Analysis based on Hashtag Research (SMAHR) framework, which uses social network analysis methods to explore social media communication through a network of hashtags. The results show that social media analysis based on hashtags provides information applicable to theoretical research and practical strategic marketing and management applications.

Keywords: social media analysis; social network analysis; hashtag; community analysis

1. Introduction

Social networks have become a common part of many people's daily lives and have become a space where individual users share information [1]. Currently, there are about 3.6 billion users of social networks, and according to forecasts, more than 4.41 billion users will use social networks in 2025 [2]. When comparing the prediction of the number of social network users and the global population for 2025 (8,184,437,460 inhabitants) [3], social networks will be used by about 54% of the world population.

In contrast to traditional mass media, such as television, newspapers and radio, content consumers have become co-creators of communication on individual social networks such as Facebook, Twitter and Instagram [4,5], in addition to private messages, of which users exchange approximately 60 billion every day on social networks. These users create over 500 million tweets per day on Twitter, upload over 95 million photos per day to Instagram, and update more than a billion statuses, post over 1.8 billion comments and upload about 480 million photos per day to Facebook [6]. Following these data, which users create from their interactions, social media analysis is a fast-growing research area aimed at extracting useful information [7], which has the potential to grow, in terms of both the number of users on social networks [8] and the growth of the digital footprint [9].

These data provide a potential for analytical researchers with many different research questions to identify users' behaviours, attitudes or experiences [10,11] and other aspects of users' lives that they project on social networks, in areas such as farmers' markets [12], organic food [13], corporate social responsibility [14], sustainability [15–17], public bike-sharing [18], environmental quality [19], gamification [20], online learning during periods

of lockdown [21], COVID-19 [22], vaccine hesitancy (monitoring vaccine-related conversations on social media) [23], dissemination of misinformation regarding COVID-19 [24], immigration [25] or public emotions about cancer [26]. As can be seen from the previous examples, social media analysis has multidisciplinary significance.

Compared to different data collection techniques, such as questionnaires, interviews or focus groups, social media analysis works with a larger sample size, for example, over 414,000 tweets [15] or over 1.3 million Instagram posts [13]. Given the integration of social networking into the daily lives of many people [27], it cannot be overlooked from a research perspective, and methodologies need to be developed to provide researchers with adequate tools for these analyses to expand knowledge, to create research questions for future qualitative research based on social media analysis and to increase validity via triangulation, using social media analysis as an alternative research method [28].

In 2018, Ideya analyzed more than 150 social technologies and services worldwide. The study found 157 existing social media monitoring tools, of which only six are free [29]. Free tools usually offer only basic characteristics, such as word frequency and the most common connections [30]. The Anson study (2017) [31] focused on the main barriers to gaining knowledge from social media by researchers, citizens, public authorities, and humanitarian organizations. The main barriers identified were the price of social media analysis tools, the absence of guidelines and frameworks for using social media data, and a lack of any acknowledgment about the value of analysing social media data.

Given the current situation, this study aimed to describe the social media analysis based on the hashtag research (SMAHR) framework. The SMAHR is a free tool for researchers to perform social media analysis that is based on social network analysis methods, and uses hashtags as a specific aspect of communication. We link our findings with those of previous studies, and present the practical implications for strategic marketing and management in the private sphere, and for policy-related areas in the field of state administration.

According to one theoretical concept, a social network is a means by which to connect with other users in a digital environment, and a social medium is a platform for mass information sharing [32].

2. Theoretical Background

Nowadays, synonymous terms for social networks are heavily intertwined and interchangeable, which can be seen in relation to Facebook, as indicated in references [33,34], where it is a social medium platform, and in references [35,36], where it is a social network. Facebook is an example of an overlap between the two areas. Reference [37] defines social networks as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system." Reference [38] defines social media as "Internet-based channels that allow users to interact opportunistically and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others."

In connection with these concepts, the term social network will be used in this article to describe a platform that enables a connection between individual users and social media platforms as possibilities of this network in the field of mass information sharing. We follow the research [39], which defines online social networking as a process whose goal is to allow people to connect with each other. Social networks are tools for sharing content [40].

More on the issue of defining the difference between social media and social network can be found in the research of Rhee (2021) [41], which focuses only on this issue.

Social Network Analysis vs. Social Media Analysis

Following the definition of the terms social network and social media, it is also necessary to define the difference between the terms of social network analysis (SNA) and social media analysis (SMA).

Social network analysis examines social structures using network theory and graph theory [42]. The importance of social network analysis can be seen in the number of articles with a title containing the keywords "Social Network Analysis" in the ScienceDirect database, which increased from 251 articles in 2010 to 1259 in 2020.

Social media analysis aims to collect, monitor, analyse, summarise, and visualise data from social media, from which useful patterns and information can be identified [43]. It is a process of collecting data from communication on digital media and processing them into structured views, which leads to better business decisions based on the obtained information [44]. Reference [45] defines social media analysis as a process of obtaining unstructured data from social media and analysing these data for areas of business decision-making based on the information obtained. This analysis is useful for product development, product innovation, product utilisation, brand engagement, competitive intelligence, and general marketing [46].

3. Social Media Analysis Based on Hashtag Research Framework

The SMAHR framework is based on the theoretical process, knowledge discovery in database (KDD) [47]. The KDD process comprises seven basic activities (see Figure 1). As with the KDD process, the SMAHR framework does not include the process of obtaining data from social media. For this part, it is necessary to use specialized software, such as Netlytic [48], or obtain data from social media through the API of individual social media.
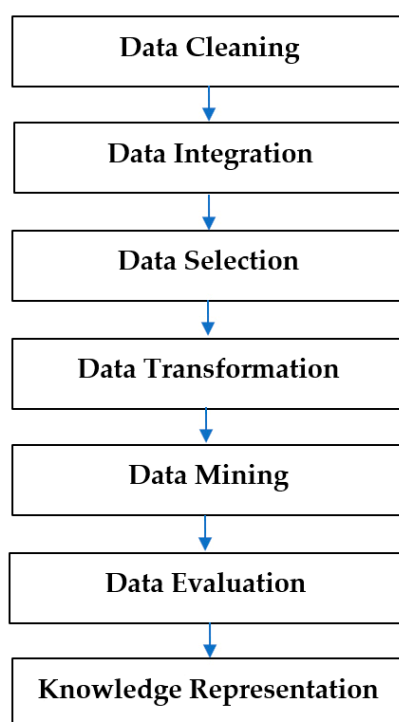

Figure 1. Seven phases of social media analysis based on hashtag research (SMAHR) framework process. Source: Own processing.

3.1. Seven Phases of the SMAHR Framework

1. **Data Cleaning**—removing irrelevant, confusing or misleading data from a data file [49]. At this stage, it is necessary to understand the context of the hashtag and the various situations for its use. It is mainly a matter of removing irrelevant and

confusing data. In this area, these are mainly hashtags that can be used in various contexts. An example is reference [14], which focused on corporate social responsibility. Netlytic software [48] was used to download the data, where a condition was specified for downloading messages that contain the hashtag #CSR (no font size limitation). Total downloads of 1,172,868 Instagram posts. Based on the analysis of communities (see phase 5—Data Mining), the community dealing with the computer game "CSR Racing" was extracted, where the hashtag #CSR was also used. This point is critical to the scope of data mining, and non-topic messages must be removed.

2. **Data Integration**—a process that is defined as data homogenization [50]. An essence of the data integration activity lies in this process in the creation of a homogeneous data set, with which it is possible to go to the third phase—Data Selection. Due to the different API settings of individual social networks, it is necessary to create a board structure for further work, containing the following items:

    - ID—identifier of an item in a data file. It is not necessary for further processing, but speeds up an orientation in an overall data set;
    - Author—this is the author's identifier of particular message. This information is important in terms of identifying a number of unique users in the data file. For example, if data have been downloaded throughout the year, one user will likely comment on the topic more than once;
    - Message—this item contains the text part of the message. If it is possible to insert hashtags into other parts of the message on a social network (some social networks have a separate message title and message body), it is necessary to merge these parts into one item;
    - Location—this is the identification of the place where the message was sent or the user's residence. This information can be used to identify regional differences (more in phase 5—Data Mining). This is a very important piece of information that can give us information whether the messages are captured throughout the period, and there is no particular outage in the data file. It can also inform us about seasonality in certain regions, which is useful, for example, in the field of farmers' markets, where in central Europe, for example, customers will communicate in connection with farmers' markets in spring other products than those produced in summer and fall [12].

3. **Data selection**—a process during which a researcher decides what data are relevant for further analysis. The SMAHR framework focuses on hashtag analysis. For this reason, it is necessary to remove all text that is not a hashtag from the message field (see the Data Integration phase) at this point. It removes any text that does not start with a "#". For this, it is possible to use the Hashtag Matcher 1.2 module, which is described in more detail in the next phase—Data Transformation. These two phases are closely related in this framework.

    Message sending location Research can be divided into two parts:

    - *An analysis of communication from a global perspective*—without regional differences. All messages are used.
    - *An analysis of a particular region, or an analysis of regional differences*—if a study is focused on identifying a regional difference or is focused on an analysis of a specific region. In this case, it is necessary to use data filtering based on values in the location field (see phase Data Integration) to select only those data that contain information about the location of the region.

An example is reference [14], which focused on identifying differences in corporate social responsibility based on the selected hashtag #CSR between developed and developing countries. In this study, 24,339 messages (21.42%) were selected from the basic set of 113,628 Instagram posts, which contain information about the message's location. A total of 9712 messages were sent from developed countries and 14,627 from developing

countries. Based on the previous step (Data Integration), this filtering can be applied, for example, in MS Excel, Apache OpenOffice, or another spreadsheet program.

4. **Data Transformation**—this is about transforming data into a suitable form, which is required by the information mining system. Basically, it is about preparing data for an information system working with data. The method of data transformation depends on the software used for data mining. The SMAHR framework recommends Gephi [51] in the current version 0.9.2. This software contains its module for importing data that are saved in CSV format. If a study is focused on regional differences or on an analysis of a specific region, it is necessary to perform data filtering based on the selected region. For this step, it is necessary to transform the data into a form where it will be possible to identify the selected region. Here, it depends very much on the API of individual social networks. At the API base, they either provide textual location information in address format or express that position through latitude and longitude, i.e., geographic coordinates. The next significant step in this phase is to transform the rest of the message into a format for data mining software. Here, it is possible to use Hashtag Matcher 1.2 software (see Supplementary Materials to this article). This software has two primary functions:

- Remove non-hashtag text (#)
- Modify hashtags in a message to the form needed for Gephi program converts:
  - ○ All text in lowercase. Hashtag #prague and #Prague are two different hashtags for Gephi. For this reason, the first part of the transformation is focused on converting all characters to lowercase;
  - ○ If two or more hashtags are connected, for example, # farmersmarket # organicfood # fresh, these hashtags are separated by a space on #farmersmarket #organicfood #fresh so that the program detects three hashtags.
  - ○ For a large data file, hashtags that are less than the specified value in the entire data file are removed. This mainly involves removal of typos, which removes all hashtags that are in the data file $1\times$.

Parameters for Hashtag Matcher 1.2:

-i—Input file.
-o—Output file. If omitted, the file name is derived from input and placed in current directory.
-l—Minimum number of occurrences for a hashtag to be included in output. Default: 0
–csv-data-column—First based index of column with data. 0 means single line. Default: 1
–csv-no-header—There is no header in the csv file. Default: false
–csv-separator—Column's separator. Default:
-m—Name of line matcher.
-h—Help.

Example: java -jar java-parserv3.jar -i = inputfile.csv -o = outputfile.csv -l = 100 -m = HashtagsMatcher –csv-data-column = 0

Based on the previous steps of Data Integration and Data Transformation, the data can be imported into a data mining process software. The essence of the data import is in the creation of a network of hashtags, based on their relationship in a given message and between hashtags in the entire data file, as seen in Figure 2. The weight value corresponds to a number of links between the given hashtags.
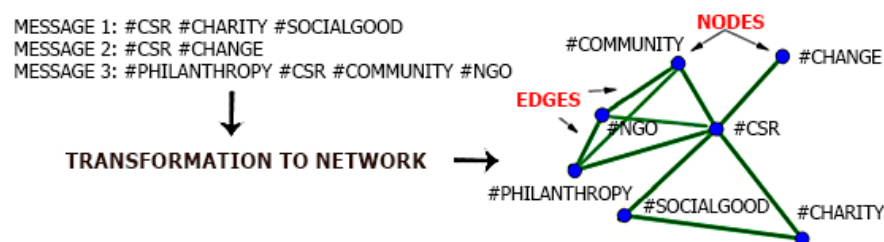
Figure 2. Transformation example of hashtags from messages to the network (case study of corporate social responsibility). Source: reference [14].

Language homogenization

If a study focuses on global communication or on conferences between regions, it is necessary to homogenize languages into one selected language.

An example is the case study of #zerowaste: it is necessary to convert the hashtag #bhayplastik (Hindi) to #bioplastic (English), and the hashtag #lebensmittelverschwendung (German) to #foodwaste (English).

5.　**Data Mining**—specific techniques used to extract potentially useful patterns. The SMAHR framework is focused on the Gephi program (currently version 0.9.2). Within this framework, the data mining techniques are used in this software in three basic areas:

- Analysis of characteristics at the level of individual hashtags;
- Analysis of network characteristics;
- Network visualization.

Analysis of characteristics at the level of individual hashtags

The following methods can be used to analyse characteristics at the level of individual hashtags:

Degree Centrality

In the field of graph theory and social network analysis, this characteristic is also called the degree of the peak. It is a degree centrality based on the importance of a vertex based on its degree in the network, which is calculated as a number of interactions with other vertices using edges. Here, it is necessary to distinguish whether the graph is oriented or non-oriented. In the area of hashtag analysis in the SMAHR framework, we use an unoriented graph (it is not possible to determine the direction of interaction between individual hashtags in the report).

Eigenvector Centrality

Centrality, according to the eigenvector, is an extension of the basic degree centrality. It is also a measure of the influence of the hashtag in the network. This value is calculated on the assumption that the connection with high-value hashtags (high-level hashtags) is more important than the connection with equally valuable or less valuable hashtags. A high eigenvector centrality score means that a hashtag is linked to many hashtags with a high value, and the formula is as follows:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t,$$

$$(1)$$

where *M (v)* denotes a set of adjacent nodes and λ is the largest eigenvalue. Eigenvector x can be expressed by Equation (2):

$$Ax = \lambda x. \tag{2}$$

Betweenness Centrality

The value of this centrality is highest for a hashtag if the paths between any two hashtags in the network always pass through this hashtag. Hashtags with a high degree of this centrality can be referred to as network bottlenecks [52]. These hashtags are important in the network because they act as interconnectors or otherwise as bridges between remote parts of the network. The suitability of this metric is in networks with a low value of modularity, where individual communities in the network are less interconnected than individual hashtags within a given community. It is then possible to identify hashtags, which are bridges between these communities. The value of the centrality for the hashtag v in the graph G : = (V, E) is calculated using the following relation:

$$C_B = \frac{\sigma_{st}(v)}{\sigma_{st}},$$

(3)

where σ_{st} is the number of shortest paths from node s to node t, and $\sigma_{st}(v_i)$ is the number of shortest paths from node s to t that pass through the node v_i.

Modularity

Most complex networks incorporate nodes that are mutually interconnected to a larger extent than with the rest of the network. Groups of such nodes are called communities [53]. Modularity represents an index that identifies the cohesion of communities within a particular network [54]. The idea is to identify node communities that are mutually interconnected to a greater degree than the others. Networks with high modularity demonstrate strong links between nodes inside modules, but weaker links between nodes in different modules [55]. The component analysis then identifies the amount of different parts (in the case of community modularity) in the network based on the modularity detection analysis [56], as follows:

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 \right],$$

(4)

where $\sum_{in}$ is the sum of weighted links inside the community, $\sum_{tot}$ is the sum of weighted links incident to hashtags in the community, k_i is the sum of weighted links incident to hashtag i, $k_{i,in}$ is the sum of weighted links going from i to hashtags in the community, and m is the normalizing factor as the sum of weighted links for the whole graph.

Network Visualization

The goal of network visualization is to identify individual communities and their relative position. After importing the data into the Gephi program, the network's visualization is concentrated in the basic square, without visualizing the different relationships of individual hashtags, as seen in Figure 3. This visualization is unsatisfactory in terms of the identification of communities and their mutual position, but for the analysis of characteristics at the level of hashtags and the whole network, there is no effect.

Based on the modularity analysis, individual hashtags are assigned colours corresponding to individual groups, which are extracted by this algorithm, as seen in Figure 4. It is now possible to determine the number of communities and their representation through component analysis. In terms of identifying a location of individual communities, this visualization is still insufficient.
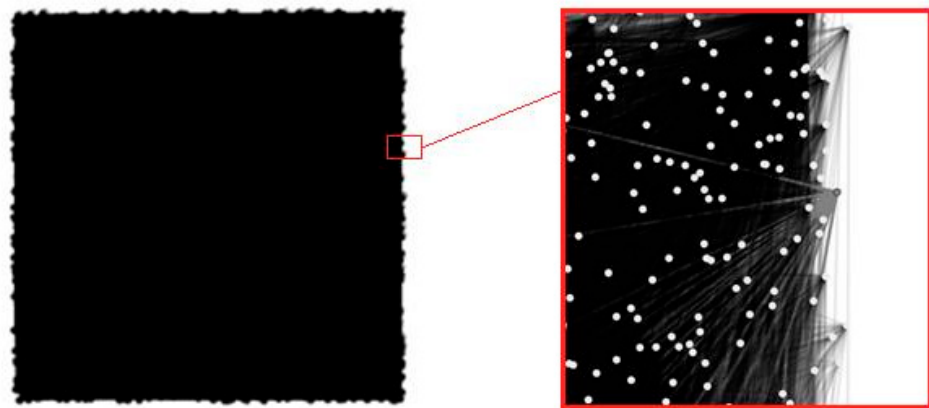
Figure 3. Default network visualization after import into Gephi (14,788 hashtags and 363,525 edges). Source: Own processing.
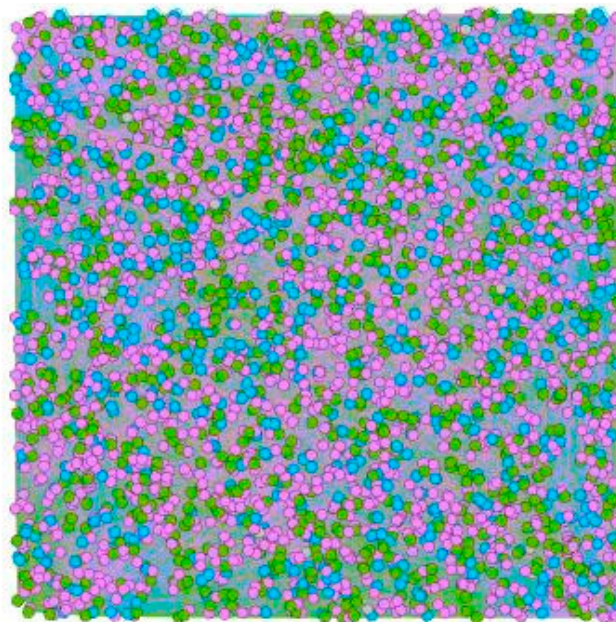

Figure 4. Network visualization in Gephi after modularity calculation without application of algorithm for layout visualization (14,788 hashtags and 363,525 edges). Source: Own processing.

In the field of visualization, it is possible to use the following algorithms, which are part of programs for visualization and analysis of data in networks: ForceAtlas, an algorithm used for networks based on the so-called small-world network. This is a type of mathematical graph in which most nodes are not neighbours, but neighbours of any given node are likely to be neighbours of each other, and most nodes can be reached from every other node by a small number of connections [57]; ForceAtlas2, an improved version of the ForceAtlas algorithm, which focuses on large networks. It is a method based on reduced samples' visual representation to define network communities and their types [58]. The advantage over ForceAtlas is its speed and ease of computing. The ideal number of hashtags is 10,000–100,000 [59]. To visualize a network with a number of hashtags from 1000 to 10,000, it is best to use the ForceAtlas 2 algorithm due to the calculation speed. The result of the ForceAtlas 2 algorithm is, for example, a subsequent layout (see Figure 5), where it is possible to identify five main communities and their mutual position (polarity).
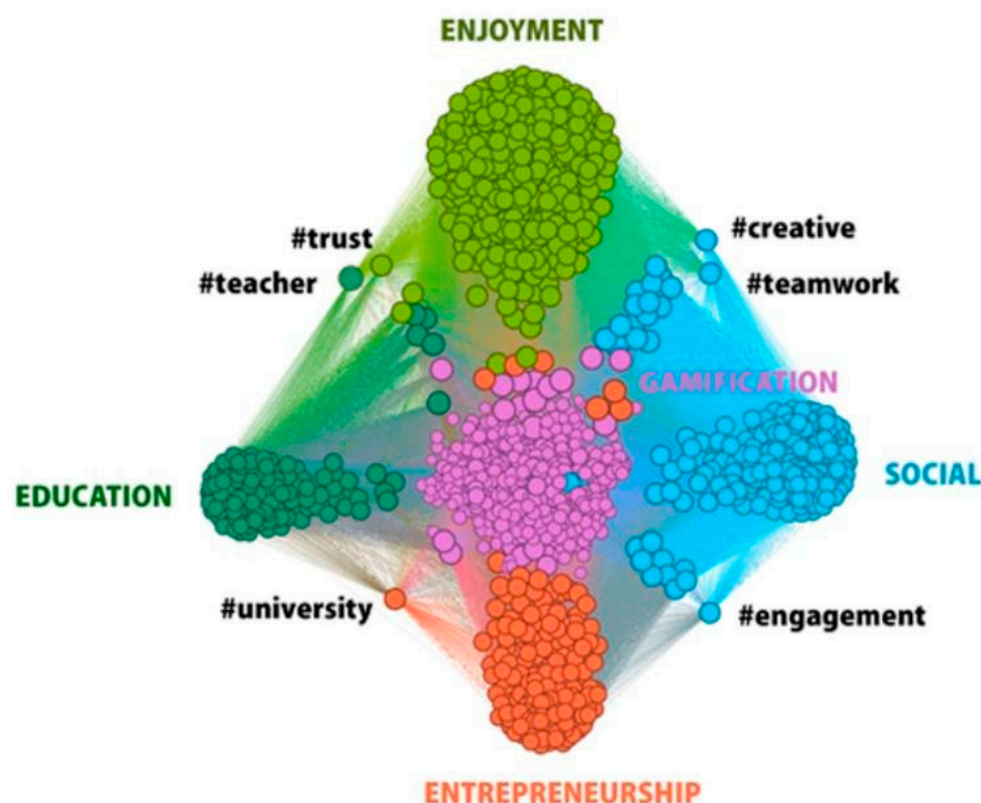
Figure 5. Network hashtag layout based on the ForceAtlas 2 algorithm (1425 hashtags and 195,993 edges). Source: "Education and Business as a Key Topics at the Instagram Posts in the Area of Gamification" [20].

6. **Data Evaluation**—the main aim is to identify specific patterns that represent knowledge based on the previous steps obtained. This is especially important in this area on the interpretation of values of the degree of hashtag, eigenvector centrality, modularity and visual assessment of the distribution of communities in the network.

Visual Assessment of the Layout of Communities in the Network

Based on identifying the modularity value, the hashtag degree values and visualizing the network, community identification is performed, where it is possible to identify communities characterizing a reduced sample. For this process, it is possible to use a division into three basic types of communities [58]. The identification of these communities was based on communication and forwarding messages between members. This theory contains six basic types of community. However, in the hashtag analysis, it is not possible to determine the hashtag's direction compared to other hashtags in the given message. Thus, they are selected and only two are described (see Figure 6), which are not conditioned by the direction of the relationship between the individual communities. These communities are made up of people who have replaced hashtags in the framework.
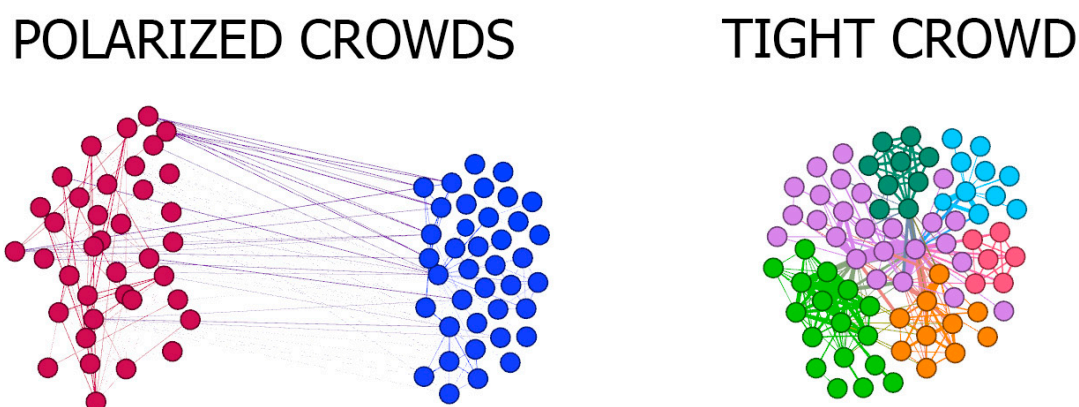
Figure 6. Basic communities of users on social networks. Adapted from: "Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters" [58].

Divided Communities—polarized communities are created for situations where two groups discuss the same topic, but communication between the topics is low. An example is reference [20], which focused on the topic of gamification, and five communities were extracted here, as seen in Figure 5.

Tight Communities—these communities can be found, for example, at conferences, discussions on professional topics and other interest groups that attract similar communities. There are different groups of conversations in these networks, but people are closely linked, even between groups. An example is reference [12], which examines farmers' markets. Four communities were extracted that are very strongly interconnected (see Figure 7).
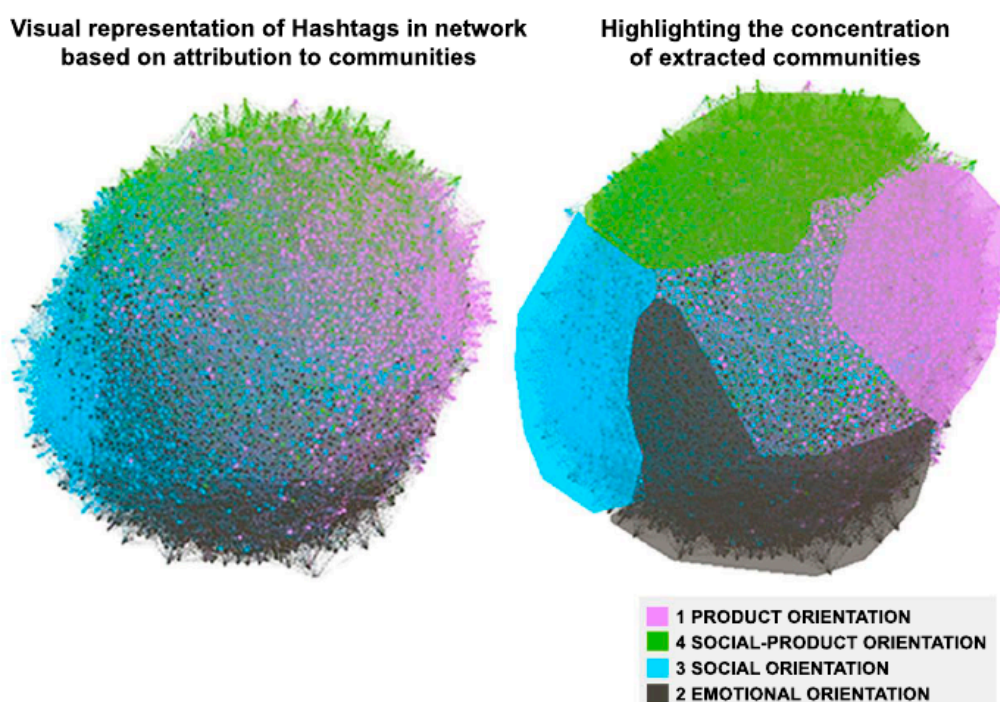

Figure 7. Example of tight communities. Source: "Customer Experience with Farmers' Markets: What Hashtags Can Reveal" [12].

The same type of community has been identified in research focusing on organic food [13].

7. **Knowledge Representation**—a technique that uses visualization tools to represent the results of data mining. Knowledge representation is based on the synthesis of

individual values and outputs from the data evaluation phase. It is possible to answer research questions here. Based on the identification of network size, network density, frequency and centralities in individual communities, it is possible to identify the essence of the network and individual communities (community focus in terms of communication) and in the context of analysis of the visual layout of the community in the network to determine the relationship of individual communities. Research questions for qualitative research should be posed in terms of the reason for this layout (polarization).

3.2. Limitation of Framework

In the field of knowledge representation, it is necessary to draw attention to the limits of this method:

Regional Differences—for an analysis that focuses on identifying attitudes without regional differences, see references [13,54]. It is necessary to draw attention to this field in terms of research limitation. Regional differences mainly affect research in two areas of a disproportionate number of reports. An example is reference [14], which identified differences in corporate social responsibility based on the selected hashtag #CSR between developed and developing countries. In this study, 24,339 messages (21.42%) were selected from a basic set of 113,628 Instagram posts, which contain information about the location of the message. Of these, 9712 messages were sent from developed countries and 14,627 from developing countries. If the analysis was created from a global perspective, the results would be affected by more communications from the region of developing countries, which are about 50% more than from developed countries. See Table 1.

Focus only on hashtags—the whole SMAHR framework is focused only on analysis through a specific part of the content, i.e., hashtags. Hashtags do not have to fully cover the meaning of the message, but they are used to express or emphasize something that is not expressed in the text of the message, such as emotions, mood, location, and/or political inclination [60]. Analyses that focus on text content without hashtag analysis have the same limitations. When interpreting the results, it is necessary to draw attention to this limitation in terms of research limitation.

Selected social network—each social network has different target users. There is a difference in users' age structure, the use of technologies by which users connect (mobile vs. desktop), education, etc. These factors affect the resulting attitudes of users.

4. Discussion

Based on the literature, we found that social media analysis can be used as a method to identify attitudes and experiences of social media users within a given area [12–14,20]. The SMAHR framework focuses only on hashtags, which is

Table 1. Regional differences vs. global view.

Developed Countries		Developing Countries		Global View	
Hashtag	**F ***	**Hashtag**	**F**	**Hashtag**	**F**
#charity	1198	#charity	1253	#charity	2451
#sustainability	847	#socialgood	817	#socialgood	1415
#socialgood	598	#education	815	#corporatesocialresponsibility	1292
#philanthropy	594	#love	758	#nonprofit	1235
#corporatesocialresponsibility	592	#nonprofit	749	#volunteer	1169
#socialimpact	552	#donate	702	#sustainability	847
#fundraising	550	#corporatesocialresponsibility	700	#education	815
#community	512	#volunteer	679	#love	758
#volunteer	490	#dogood	639	#donate	702
#nonprofit	486	#ngo	631	#dogood	639

* F—frequency. Source: reference [14].

one of the limitations of this work.

The uniqueness of this framework lies mainly in the fact that it covers social media analysis but uses the methods of social network analysis. With this specific focus, the SMAHR is a competitive framework that can be used as an alternative framework in the field of social media analysis, most of which are focused primarily on semantic and sentiment analysis [30,61–64]. This thus provides a tool for research triangulation, which has already been confirmed by previous work [28].

As mentioned above, hashtags are a specific part of communication through which users express experience, attitudes, opinions, and values [30,60,65]. Thus, the user can use hashtags to express characteristics that may not be obvious from the post alone [60,66]. For this reason, hashtags are very often inserted at the end of a message (see Figure 8 for an example).

Figure 8. Example of an Instagram message with hashtags.

Figure 8 shows an example from Instagram [67], which was filtered based on the hashtag #onions. The message contains three main parts for analysis, as follows: (1) image, (2) text, and (3) hashtags.

(1) Image: Using machine learning models for the classification of image content, it is possible to detect objects (i.e., vegetables, onions, cucumber, car, building, etc.) and context (i.e., situation and association, such as natural food, celebration, etc.). In this case, the cloud vision API algorithm [68] was used. This algorithm has been used to, for example, extract brand information from social networks [69] and to recognize cultural ecosystem services from social media photographs [70].

Google Vision correctly detected vegetable and onion (see Figure 9 and Table 2), but failed to identify squash, which is not in the image. The percentage corresponds to the confidence that the algorithm assigns to the detected objects.
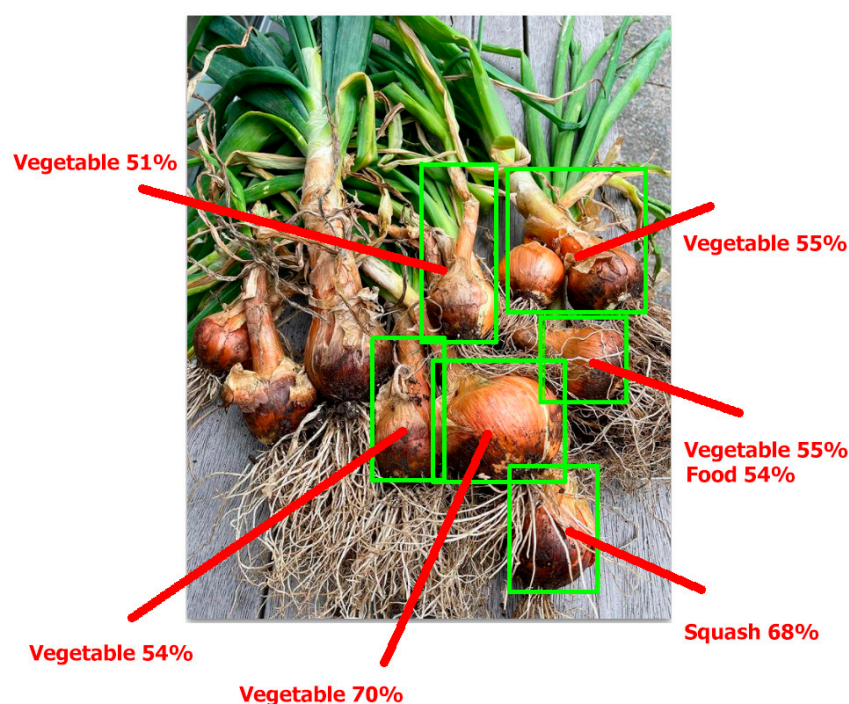
Figure 9. Output from Google Vision—Object identification.

Table 2. Labels based on Google Vision (example for an image of an onion).

Labels	Confidence	Labels	Confidence
Food	97%	Local Food	74%
Plant	93%	Whole Food	69%
Ingredient	89%	Bulb	62%
Natural Foods	88%	Allium	61%
Onion	83%	Superfood	61%
Terrestrial Plant	83%	Herb	61%
Root Vegetable	82%	Shallot	57%
Vegetable	80%	Vegan Nutrition	56%
Produce	77%	Yellow Onion	56%
Staple Food	76%	Root	56%
Flowering Plant	75%	Beet	52%

Based on the algorithm, Google Vision derived the context of the message, that is, local food (see Table 2), which was not mentioned either in the text of the message or in the hashtags. Estimates such as this are based on the analysis of similar images, which has an algorithm in the database.

(2) Text: Frameworks that focus on social media analysis using natural language processing may not identify certain kinds of information because the report from which the hashtags are removed may not contain information. In the case of an onion image (Figure 8), the Cloud Natural Language algorithm [71] was used. This algorithm has been used in thematically similar studies for sentiment analysis of police agency Facebook pages before and after a fatal officer-involved shooting of a citizen [72], and sentiment analysis of consumer reviews [73]. In the case of an onion, the algorithm identified one category "Food & Drink/Food" with a confidence of 52%. In the area of sentiment analysis, the algorithm identified the message as neutral, with a value of 0.1 (the score of the sentiment ranges from -1.0 (very negative) to 1.0 (very positive)).

(3) Hashtags: The message contained the following 22 hashtags: #kitchengardenz #onions #vegegarden #kitchengarden #gardentotable #raisedbeds #growingvegetables #growingfood #urbanpermaculture #foodforest #organicgarden #selfsufficiency #sustain-

ablelifestyle #homestead #ediblegarden #raisedbedgardening #permaculture #veggiepatch #homegrownfood #nzblogger #gardensofinstagram #gardeningnz #myprideofplace #myawapunigarden.

In the case of the onion, a high-quality photo was chosen. Social network users do not create posts following our opportunity to analyse them. This means that the photo may not be of a high quality, and the algorithm for detecting objects in the photo may misidentify objects. There is no condition for creating messages on Instagram to insert text, so an algorithm with a focus on natural language processing cannot be used. Thus, this limitation of individual social networks also limits the use of algorithms that apply natural language processing. For example, Twitter limits users to a maximum post length of 280 characters. This means that Twitter users often express themselves in a different way than they would naturally in everyday life because, in most communication scenarios, one is not limited to 280 characters.

Based on the comparison of machine learning models for the classification of an image, natural language processing, and hashtag analysis, it is not possible to determine a better method for mass processing of communication on social networks. However, to gain knowledge, it is necessary to use triangulation processes in the form of a mutual combination of these methods, which can be supported by qualitative research such as questionnaires [28].

Practical Implications

Previous work has found that the SMAHR framework has practical implications. Research [12] focused on farmer's markets identified four customer segments and recommended that sellers offer a diverse product portfolio at farmer's markets, considering the characteristics of identified customer segments. These results are in accordance with two previous studies [60,61], which collected data from the farmers' markets in Beijing, especially in Gipuzkoa. The study identified six communities. Knowledge about these communities and their polarization can give organizations insights into possible directions for improving their sustainable business model via the contribution of public opinion. One study [14] compared CSR communication in developing and developed countries, and showed that there are different types of communication on social networks; the authors recommend that global companies do not create a unified communication strategy, but rather that they adjust the strategy for each particular region. This approach has also been supported by one study [63] that reported findings from two online surveys. The SMAHR framework has also been used to identify the main topic associated with healthy food on Twitter [74]. That study found that people on social networks connect food with lifestyle, which is not just about nutrition, but about the way a person lives. Future research could focus on the prediction of communication on social networks using artificial intelligence [75], using hashtags, as well as the identification of topic changes over time using structural topic modelling [76].

5. Conclusions

Social media has become a part of many people's lives, and the analysis of data that users create through their interaction with social networks is thus becoming a very current topic for many areas of research. This article confirms that hashtags, as a specific part of communication, provide a research opportunity to identify the attitudes and experience of users in social networks in any given area. Moreover, the SMAHR framework provides a methodological procedure and software module for working with hashtags.

The presented framework systematically codifies the procedure into a few main steps which could be adapted for future research. By using this framework, researchers can focus on their research goals and resolve typical problems in the analysis process (such as data selection and transformation). The steps described above will help researchers to process all the actions to clarify the results. This is a powerful tool that combines quantitative analysis of hashtags with other research.

As a tool for hashtag analysis, the SMAHR framework will help identify clusters in knowledge analysis of the social network. It will help to categorize contextual information contained in posts, provide an opportunity to categorize user groups based on relevant criteria.

The results in the case studies discussed in this article show that social media analysis based on social network analysis methods which focus on hashtags provides information for theoretical application in scientific research and practical application in strategic marketing and management in many research areas (e.g., organic food, farmer's markets, gamification, sustainability). Some approaches and possibilities for using the framework are suggested in our discussion as model situations for future research.

Social analysis provides researchers with diverse data and multidimensional information. Based on the different analyses of content, we can explain many of the actual problems in different fields. Comparing hashtag analyses will help create a comprehensive, fruitful picture of users' posts.

References

1. Wu, J.; Wang, Y. *Opportunistic Mobile Social Networks*; CRC Press: Boca Raton, FL, USA, 2014; ISBN 9781466594951.
2. Clement, J. Number of Social Network Users Worldwide from 2017 to 2025. Available online: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (accessed on 14 January 2020).
3. Worldometer World Population Projections. Available online: https://www.worldometers.info/world-population/world-population-projections/ (accessed on 4 February 2021).
4. Apuke, O.D. Social and traditional mainstream media of communication: Synergy and variance perspective. *Online J. Commun. Media Technol.* **2017**, *7*. [CrossRef]
5. Malthouse, E.C.; Haenlein, M.; Skiera, B.; Wege, E.; Zhang, M. Managing customer relationships in the social media era: Introducing the social crm house. *J. Interact. Mark.* **2013**, *27*, 270–280. [CrossRef]
6. Smith, K. 126 Amazing Social Media Statistics and Facts. Available online: https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/ (accessed on 2 January 2021).
7. Cesario, E.; Marozzo, F.; Talia, D.; Trunfio, P. SMA4TD: A Social Media Analysis Methodology for Trajectory Discovery in Large-Scale Events. *Online Soc. Netw. Media* **2017**, *3–4*, 49–62. [CrossRef]
8. Asur, S.; Huberman, B.A.; Szabo, G.; Wang, C. Trends in Social Media: Persistence and Decay. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
9. Mou, N.; Zheng, Y.; Makkonen, T.; Yang, T.; Tang, J.; Song, Y. Tourists' digital footprint: The spatial patterns of tourist flows in Qingdao, China. *Tour. Manag.* **2020**, *81*, 104151. [CrossRef]

10. Zhang, K.; Geng, Y.; Zhao, J.; Liu, J.; Li, W. Sentiment analysis of social media via multimodal feature fusion. *Symmetry* **2020**, *12*, 2010. [CrossRef]

11. Gon, M. Local experiences on instagram: Social media data as source of evidence for experience design. *J. Destin. Mark. Manag.* **2020**, *19*, 100435. [CrossRef]

12. Pilař, L.; Balcarová, T.; Rojík, S.; Tichá, I.; Poláková, J. Customer experience with farmers' markets: What hashtags can reveal. *Int. Food Agribus. Manag. Rev.* **2018**, *21*, 755–770. [CrossRef]

13. Pilař, L.; Stanislavská, L.K.; Poláková, J.; Rojík, S.; Kvasnička, R.; Gresham, G. Customer experience with organic food: Global view. *Emir. J. Food Agric.* **2018**. [CrossRef]

14. Stanislavská, L.K.; Pilař, L.; Margarisová, K.; Kvasnička, R. Corporate social responsibility and social media: Comparison between developing and developed countries. *Sustainability* **2020**, *12*, 5255. [CrossRef]

15. Pilař, L.; Stanislavská, L.K.; Pitrová, J.; Krejčí, I.; Tichá, I.; Chalupová, M. Twitter analysis of global communication in the field of sustainability. *Sustainability* **2019**, *11*, 6958. [CrossRef]

16. Domalewska, D. A longitudinal analysis of the creation of environmental identity and attitudes towards energy sustainability using the framework of identity theory and big data analysis. *Energies* **2021**, *14*, 647. [CrossRef]

17. Lee, J.H.; Wood, J.; Kim, J. Tracing the trends in sustainability and social media research using topic modeling. *Sustainability* **2021**, *13*, 1269. [CrossRef]

18. Serna, A.; Ruiz, T.; Gerrikagoitia, J.; Arroyo, R. Identification of enablers and barriers for public bike share system adoption using social media and statistical models. *Sustainability* **2019**, *11*, 6259. [CrossRef]

19. Wang, Z.; Ke, L.; Cui, X.; Yin, Q.; Liao, L.; Gao, L.; Wang, Z. Monitoring environmental quality by sniffing social media. *Sustainability* **2017**, *9*, 85. [CrossRef]

20. Pilař, L.; Moulis, P.; Pitrová, J.; Bouda, P.; Gresham, G.; Balcarová, T.; Rojík, S. Education and Business as a Key Topics at the Instagram Posts in the Area of Gamification. *J. Effic. Responsib. Educ. Sci.* **2019**, *12*, 26–33. [CrossRef]

21. Harper, F.K.; Rosenberg, J.M.; Comperry, S.; Howell, K.; Womble, S. #Mathathome during the COVID-19 Pandemic: Exploring and reimagining resources and social supports for parents. *Educ. Sci.* **2021**, *11*, 60. [CrossRef]

22. Chen, N.; Zhong, Z.; Pang, J. An exploratory study of COVID-19 information on twitter in the greater region. *Big Data Cogn. Comput.* **2021**, *5*, 5. [CrossRef]

23. Piedrahita-Valdés, H.; Piedrahita-Castillo, D.; Bermejo-Higuera, J.; Guillem-Saiz, P.; Bermejo-Higuera, J.R.; Guillem-Saiz, J.; Sicilia-Montalvo, J.A.; Machío-Regidor, F. Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines* **2021**, *9*, 28. [CrossRef] [PubMed]

24. Wani, M.A.; Agarwal, N.; Bours, P. Impact of unreliable content on social media users during COVID-19 and stance detection system. *Electronics* **2020**, *10*, 5. [CrossRef]

25. Vilella, S.; Lai, M.; Paolotti, D.; Ruffo, G. Immigration as a Divisive Topic: Clusters and Content Diffusion in the Italian Twitter Debate. *Futur. Internet* **2020**, *12*, 173. [CrossRef]

26. Park, S.K.; Park, H.-A.; Lee, J. Understanding the Public's emotions about cancer: Analysis of social media data. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7160. [CrossRef]

27. Hong, J.; Sila-Nowicka, K.; McArthur, D.P. Is the popularity of social networking services beneficial for public health? Focusing on active travel and BMI. *J. Transp. Health* **2018**, *11*, 183–192. [CrossRef]

28. Pilař, L.; Stanislavská, L.K.; Gresham, G.; Poláková, J.; Rojík, S.; Petkov, R. Questionnaire vs. Social Media Analysis—Case Study of Organic Food. *Agris Online Pap. Econ. Inform.* **2018**, *10*, 93–101. [CrossRef]

29. IDEYA Social Media Monitoring Tools and Services Report Public Excerpts 2018. Available online: http://ideya.eu.com/images/SMMTools%20Excerpts%202018%20Final.pdf (accessed on 4 February 2021).

30. Maynard, D.; Roberts, I.; Greenwood, M.A.; Rout, D.; Bontcheva, K. A framework for real-time semantic social media analysis. *J. Web Semant.* **2017**, *44*, 75–88. [CrossRef]

31. Anson, S.; Watson, H.; Wadhwa, K.; Metz, K. Analysing social media data for disaster preparedness: Understanding the opportunities and barriers faced by humanitarian actors. *Int. J. Disaster Risk Reduct.* **2017**, *21*, 131–139. [CrossRef]

32. Vlad, D.E. Concepts of quality connected to social media and emotions. In *Sustainable Management, Wertschöpfung und Effizienz*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2020; ISBN 978-3-658-28866-2.

33. Ruggieri, S.; Santoro, G.; Pace, U.; Passanisi, A.; Schimmenti, A. Problematic facebook use and anxiety concerning use of social media in mothers and their offspring: An actor–partner interdependence model. *Addict. Behav. Reports* **2020**, *11*, 100256. [CrossRef]

34. Seidel, R.L.; Jalilvand, A.; Kunjummen, J.; Gilliland, L.; Duszak, R. Radiologists and social media: Do not forget about facebook. *J. Am. Coll. Radiol.* **2018**, *15*, 224–228. [CrossRef]

35. Chen, Y.-R.R.; Zhao, X. Digital dialogue in online brand communities: Examining the social network outcomes of brands' dialogue with facebook users. *Telemat. Inform.* **2020**, 101507. [CrossRef]

36. Berezan, O.; Krishen, A.S.; Agarwal, S.; Kachroo, P. Exploring loneliness and social networking: Recipes for hedonic well-being on facebook. *J. Bus. Res.* **2020**, *115*, 258–265. [CrossRef]

37. Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. *J. Comput. Commun.* **2007**, *13*, 210–230. [CrossRef]

38. Carr, C.T.; Hayes, R.A. Social media: Defining, developing, and divining. *Atl. J. Commun.* **2015**, *23*, 46–65. [CrossRef]

39. Das, B.; Sahoo, J.S. Social Networking Sites—A Critical Analysis of Its Impact on Personal and Social Life. *Int. J. Bus. Social Sci.* **2015**, *5*, 222–228.

40. Hur, K.; Kim, T.T.; Karatepe, O.M.; Lee, G. An exploration of the factors influencing social media continuance usage and information sharing intentions among korean travellers. *Tour. Manag.* **2017**, *63*, 170–178. [CrossRef]

41. Rhee, L.; Bayer, J.B.; Lee, D.S.; Kuru, O. Social by definition: How users define social platforms and why it matters. *Telemat. Inform.* **2021**, *59*, 101538. [CrossRef]

42. Otte, E.; Rousseau, R. Social network analysis: A powerful strategy, also for the information sciences. *J. Inf. Sci.* **2002**, *28*, 441–453. [CrossRef]

43. Zeng, D.; Chen, H.; Lusch, R.; Li, S.-H. Social media analytics and intelligence. *IEEE Intell. Syst.* **2010**, *25*, 13–16. [CrossRef]

44. Dozier, C. Social Media Analytics Market Growing Technology Opportunities and Future Business Trends to 2025. Available online: https://marketresearchupdates.com/2019/03/13/social-media-analytics-market-growing-technology-opportunities-future-business-trends-2025-ibm-oracle-salesforce-adobe-systems-sas-institute/ (accessed on 4 February 2021).

45. Stieglitz, S.; Mirbabaie, M.; Ross, B.; Neuberger, C. Social media analytics—Challenges in topic discovery, data collection, and data preparation. *Int. J. Inf. Manag.* **2018**, *39*, 156–168. [CrossRef]

46. Fan, W.; Gordon, M.D. The power of social media analytics. *Commun. ACM* **2014**, *57*, 74–81. [CrossRef]

47. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: Waltham, MA, USA, 2011; ISBN 9780123814791.

48. Gruzd, A. Netlytic: Software for Automated Text and Social Network Analysis. Available online: http://netlytic.org (accessed on 4 February 2021).

49. Nisbet, R.; Miner, G.; Elder, J. *Handbook of Statistical Analysis and Data Mining Applications*; Academic Press: London, UK, 2009; ISBN 9780123747655.

50. Gómez-Jiménez, G.; Gonzalez-Ponce, K.; Castillo-Pazos, D.J.; Madariaga-Mazon, A.; Barroso-Flores, J.; Cortes-Guzman, F.; Martinez-Mayorga, K. The OECD principles for (Q)SAR models in the context of knowledge discovery in databases (KDD). *Adv. protein Chem. Struct. Biol.* **2018**, *113*, 85–117. [PubMed]

51. Bastian, M.; HEymann, S.; Jacomy, M. Gephi: An Open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, 17–20 May 2009.

52. Mittal, R.; Bhatia, M.P.S. Discovering bottlenecks entities in multi-layer social networks. *J. Discret. Math. Sci. Cryptogr.* **2019**, *22*, 241–252. [CrossRef]

53. McCurdie, T.; Sanderson, P.; Aitken, L.M. Applying social network analysis to the examination of interruptions in healthcare. *Appl. Ergon.* **2018**, *67*, 50–60. [CrossRef] [PubMed]

54. Newman, M.E.J.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113. [CrossRef] [PubMed]

55. Knoke, D.; Yang, S. *Social Network Analysis*; Sage Publications: London, UK, 2008; Volume 154, ISBN 9781412927499.

56. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [CrossRef]

57. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440–442. [CrossRef]

58. Smith, A.; Shneiderman, B.; Himelboim, I. Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. Available online: http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/ (accessed on 4 February 2021).

59. Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE* **2014**, *9*, e98679. [CrossRef]

60. Pilař, L.; Poláková, J.; Gresham, G.; Rojík, S.; Tichá, I. Why People Use Hashtags When Visiting Farmers' Markets. In Proceedings of the Agrarian Perspectives Xxvi: Competitiveness Of European Agriculture and Food Sectors, Prague, Czech Republic, 13–15 September 2017; pp. 287–292.

61. Ali, K.; Dong, H.; Bouguettaya, A.; Erradi, A.; Hadjidj, R. Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework. In Proceedings of the 2017 IEEE International Conference on Web Services (ICWS), Honolulu, HI, USA, 25–30 June 2017; pp. 660–667.

62. AlKhatib, M.; El Barachi, M.; AleAhmad, A.; Oroumchian, F.; Shaalan, K. A sentiment reporting framework for major city events: Case study on the china-united states trade war. *J. Clean. Prod.* **2020**, *264*, 121426. [CrossRef]

63. Singh, R.K.; Verma, H.K. Effective parallel processing social media analytics framework. *J. King Saud Univ. Comput. Inf. Sci.* **2020**. [CrossRef]

64. Zhan, Y.; Han, R.; Tse, M.; Ali, M.H.; Hu, J. A social media analytic framework for improving operations and service management: A study of the retail pharmacy industry. *Technol. Forecast. Soc. Change* **2021**, *163*, 120504. [CrossRef]

65. Childers, C.C.; Lemon, L.L.; Hoy, M.G. #Sponsored #Ad: Agency perspective on influencer marketing campaigns. *J. Curr. Issues Res. Advert.* **2019**, *40*, 258–274. [CrossRef]

66. Sheldon, P.; Bryant, K. Instagram: Motives for Its Use and Relationship to Narcissism and Contextual Age. *Comput. Hum. Behav.* **2016**, *58*, 89–97. [CrossRef]

67. Thecottagecannery just Posted Onions Today. Available online: https://www.instagram.com/p/CM6N043hV62/?utm_source=ig_web_button_share_sheet (accessed on 6 April 2021).

68. Google Cloud Vision API. Available online: https://cloud.google.com/vision (accessed on 2 April 2021).
69. Klostermann, J.; Plumeyer, A.; Böger, D.; Decker, R. Extracting brand information from social networks: Integrating image, text, and social tagging data. *Int. J. Res. Mark.* **2018**, *35*, 538–556. [CrossRef]
70. Richards, D.R.; Tunçer, B. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosyst. Serv.* **2018**, *31*, 318–325. [CrossRef]
71. Google Natural Language. Available online: https://cloud.google.com/natural-language (accessed on 3 April 2021).
72. Hand, L.C.; Ching, B.D. Maintaining neutrality: A sentiment analysis of police agency facebook pages before and after a fatal officer-involved shooting of a citizen. *Gov. Inf. Q.* **2020**, *37*, 101420. [CrossRef]
73. Al-Natour, S.; Turetken, O. A comparative assessment of sentiment analysis and star ratings for consumer reviews. *Int. J. Inf. Manag.* **2020**, *54*, 102132. [CrossRef]
74. Pilař, L.; Stanislavská, L.K.; Kvasnička, R. Healthy food on the twitter social network: Vegan, homemade, and organic food. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3815. [CrossRef]
75. Camacho, D.; Luzón, M.V.; Cambria, E. New Research Methods & Algorithms in Social Network Analysis. *Futur. Gener. Comput. Syst.* **2021**, *114*, 290–293. [CrossRef]
76. Fridrich, M. Understanding customer churn prediction research with structural topic models. *Econ. Comput. Econ. Cybern. Stud. Res.* **2020**, *54*, 301–317. [CrossRef]