A Glimpse on Big Data Analytics in the Framework of Marketing Strategies

Pietro Ducange, Riccardo Pecori, Paolo Mezzina

Abstract Mining and analyzing the valuable knowledge hidden behind the amount of data available in social media is becoming a fundamental prerequisite for any effective and successful strategic marketing campaign. Anyway, to the best of our knowledge, a systematic analysis and review of the very recent literature according to a marketing framework is still missing. In this work, we intend to provide, first and foremost, a clear understanding of the main concepts and issues regarding social big data, as well as their features and technologies. Secondly, we focus on marketing, describing an operative methodology to get useful insights from social big data. Then, we carry out a brief but accurate classification of recent use cases from the literature, according to the decision support and the competitive advantages obtained by enterprises whenever they exploit the analytics available from social big data sources. Finally, we outline some open issues and suggestions in order to encourage further research in the field.

Keywords Social Big Data · Social Media · Social Networks · Strategic Marketing

1 Introduction

The Internet has experienced a constant growth and development, both in the past and nowadays, creating digital traces that can be collected and processed to define different individual schemes, themselves useful to discern both single-

Pietro Ducange · Riccardo Pecori
SMART Engineering Solutions & Technologies (SMARTEST) Research Centre
eCAMPUS University, Via Isimbardi 10, 22060, Novedrate CO, Italy
email:riccardo.pecori@uniecampus.it

and group-related behaviors. This is indeed a time when a huge quantity of information has an outstanding speed of diffusion throughout the networks: data equipped with these features have thereby been labeled as *big data* [31, 61]. Some figures per minute, hinting at an approximate dimension of the phenomenon, are the following: millions of e-mails, Twitter and Facebook posts, thousands of Instagram photos and the like.

As a consequence, such massive volumes of data are becoming a basic feature of the society and, at the same time, the ability to analyze, correlate and learn from them is turning into a useful element to compete as well as to support growth, productivity and innovation in different fields [11, 56]. In this perspective, broadband Internet connections and social media are playing a fundamental role in strengthening the relationship between companies and their customers: they have the power to dramatically change the marketing strategies [30] and the political attitudes making successful predictions [75]. This is also proven by the movement of a great amount of investments towards social media-driven decision systems, during the last years, to the detriment of traditional alternatives [15, 82].

Although big data can be definitely considered a blessing for decision-making, having big data does not automatically lead to better marketing as they are intertwined with some important challenges and issues: the impossibility to use a unique central unit and classical storage facilities, the need for real time analytics, the correctness of the insights, privacy preservation and so on [53], [60].

An important role in the context of social media analytics is played by machine learning and computational intelligence techniques. Indeed, text mining, user profiling and localization, sentiment analysis, social sensing and the like, are just some of the means used to perform a deep analysis of the traces that people continuosly leave on social media [55, 57, 94]. Recently, a number of contributions for expoiting machine learning and computational intelligence techniques for big data analysis have been discussed in the specialized litterature [45, 46, 65, 91].

In this paper, we want to provide some operative and technological advice on how to harness the wealth of data available on social media to enhance corporate marketing strategies. We first review the main features and issues related to social media, big data and the current technologies to deal and analyze them. Then, we offer a glimpse of an operative method, inspired by the works of Berkhout et al. [13, 14], that we suggest to exploit for social big data analytics. In this way, we provide some hints on how gathering information from unstructured data, present in the Internet, and from structured data already owned by enterprises, in order to reinforce a proactive and emotional action for the, possibly real-time, management of customers' needs. Finally, in order to highlight those issues still unsolved and shed light over new possible research paths, we also perform a brief review of prior research works and use

cases of the recent literature according to a four-level framework: integration, development, brand reputation and customer relationship.

The sections ahead of the paper are organized as follows: Sec. 2 summarizes the current main sources of big data, namely on-line social media, and provides some hints to understand the big data phenomenon, describing some intrinsic characteristics of social big data and giving a brief summary of the main advantages and drawbacks they provide to the society. In Sec. 3, we review some recent IT technologies that can be employed for handling and analyzing social big data. Moreover, we also draw an operative methodology for making social big data analytics an effective marketing tool. Sec. 4 presents an assessment and classification of use cases regarding marketing applications of analytics obtained from social big data in the perspective of highlighting what has been already investigated and future research issues. Finally, Sec. 5 sums up the article with some conclusions and possible future developments and topics for better exploiting the social big data phenomenon.

2 Society 2.0 and Social Big Data

Human beings have always tried to "datify" the world, but currently the novel information technologies allow them to strongly increase the ease and the speed of transforming each human and social phenomenon into data; as said by the authors in [61], the world is becoming more and more "datified". Indeed, the XXI century is more and more composed by a generation living, entertaining and studying [28] on the Internet that leaves digital marks, often elaborated to produce a detailed description of both individual and group behaviors. This leads to the possibility of transforming the perception of lives and of the whole society.

Society 2.0 is a neologism used to define the interconnected world we live in, vouched by the flood of big data currently present on the Internet. This is the realization of past theories [10] enacting a new way of thinking and new economic and social dependencies. Society 2.0 is based on social media, web-based means of interaction among people, through which they create, share and exchange various kinds of information ranging from pictures to texts, from music to videos and so on, in a sort of virtual community. Some of the most popular are Facebook, Twitter, LinkedIn, YouTube, Instagram, Google+, Tumblr, Flickr, emails, forums, blogs and so on. Among these, a subset, known as Online Social Networks (OSNs), is garnering a major part in the big data scenario, mainly thanks to its ability to mirror existent bonds in real life, such as those with friends, relatives, colleagues and the like [76]. The purpose of a social network may be multifaceted: interaction with friends and families, creation of new business contacts, sharing photos and experiences, sharing emotions and feelings at any time and across distance, meeting new people, and so forth. Anyway, they are one of the most used ways people exploit

to communicate and this is clear from Fig. 1, courtesy of [1]. It shows the number of digital data generated, on average per minute, by social media on the Internet in 2014, 2013 and 2012.
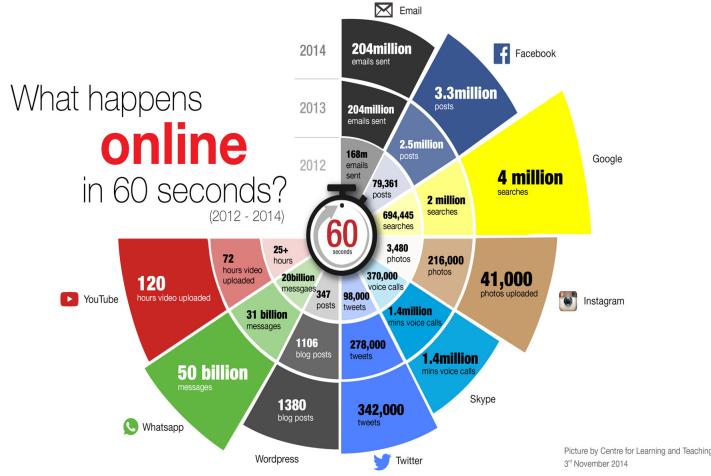

Fig. 1: Social Big Data numbers in one minute. Courtesy of CLT - Hong Kong [1].

Apart from giving a hint on the ongoing dimensions of social big data, Fig. 1 also verifies the constant and exponential increase of the circulating digital data over the years. Another emerging concept is that the more we are able to share (data, information, etc.), the more we will be able to garner; this is an idea by now present in all those people belonging to the so-called "millenials" generation, i.e., those born in the '80s and '90s, and even more in the so-called "digital natives" or "generation Z", born in the 2000s and on. Finally, as vouched for in [38], thanks to social big data, the world and the society are shrinking; this is related to what is known as "degrees of separation", i.e. the number of intermediaries between two individuals in the real world, which Milgram, in the '60s, showed to be between 4.4 and 5.7 [63]. In the era of social media and social networks, thanks to social big data and connections, these degrees drop to less than 4 [38].

2.1 The Vs of Big Data

In this subsection we perform an analysis of the features of social big data, shedding light on their evolution and on the issues and challenges connected with them.

In 2001, Laney introduced three basic concepts [51] to describe big data:

- *Volume*, referred to the huge quantity of data, but also to the fact that more than 90% of the world data were generated in the last years [62]. The issues related to this feature regard to acquire, store and rein in an efficient way even zettabytes of information of any type that, in turn, must be organized, verified and analyzed.

- *Velocity*, intended both as the relentless rapidity with which data spread throughout networks, and as the quickness needed to parse data in real-time. This is a key feature of big data and marks a difference between them and a simple large data set. A possible issue concerns the identification of a trend or an opportunity in a few minutes before competitors in order to attain a competitive advantage, as analyzing data two minutes later could be very dangerous in a vying market.

- *Variety*, considered as the different types of data coming from different sources, such as sensors, social networks, and other devices and applications, but also seen as a different kind of richness, which traditional data sets are not able to convey. The point here is that traditional technologies are perhaps unable to deal efficiently and contemporaneously with structured, semi-structured and unstructured data, and new specific solutions are needed.

Indeed, as time went by, Laney's three Vs have proven inadequate to describe big data completely, and along the years, new Vs were added in order to properly characterize them. Therefore, a shift has taken place from a three Vs model to a five Vs scheme [16] and, more recently, to a seven and nine Vs pattern [67]. The five-V model adds *Virality* and *Vector* to the framework. The latter refers to the possible directions and dimensions of the spreading of big data and their inherent geolocalization and GPS information, the former concerns the percentage of perfect temporal arrival of data and information to certain nodes or users.

However, the most recent models encompass nine Vs [67], and in the transition from the five-V model to the one featuring nine Vs, an evolution occurred in the concepts of Virality and of Vector. The first concept can be incorporated in both the Veracity and the Value notions. The second one has been absorbed partially by the Visualization idea and partially by the Variability conceit, considering the meaning of space information. Moreover, Volatility and Validity have been introduced. In the following, we list the additional six Vs, accompanying the traditional three ones, as well as some related issues:

- *Veracity*, in the sense of truthfulness and accuracy. The issues connected with this feature regard the quality and the trustworthiness of the insights to construct a novel added value. The former may impact supporting decision processes, while trustworthiness is becoming more and more important in various IT fields [68] and the ranking and dependability of the retrieved

data are becoming a fundamental prerequisite to devise new ideas and intuitions.

– *Variability.* It concerns the changes, over time and over different contexts, the connotation and significance of any given datum may experience. The related issue is to frame the insights in a particular situation to decode the exact sense of a word or sentence in specific circumstances.

– *Visualization* concerns the hard process of making huge amounts of data understandable and interpretable for better user experiences and services. This is not the most difficult task but it is surely one of the most hard and crucial.

– *Validity* regards the correct usage of uncorrupted data. It is tightly bound with Veracity but considering data integrity.

– *Volatility* encompasses the retention policy of data and the period they should be stored in for future useful usage.

– *Value.* This is the most important gift social big data could offer to firms, businesses and customers. The worthiness is in their investigation and parsing and in the way they turn into useful information, new knowledge, competitive advantage and return on the investments.

3 Technologies and Methodologies for exploiting Social Big Data for Marketing Decisions

3.1 The main technologies to handle social big data

The intersection between big data and social media is reflected in the employment, by social networks, of the big data technologies in the design and development of their IT infrastructures. As a matter of fact, classical programming and storage paradigms cannot be employed in order to handle the huge amount of data coming from social media and social networks. For this reason, Google defined and patented the distributed MapReduce programming model [27]. MapReduce is based on three elements: mapper, combiner and reducer functions. To effectively sum it up, input data are mapped to key values that are combined and grouped together, according to their similarities, and finally reduced to a minimum set of output values [27]. The MapReduce paradigm may be implemented within different programming frameworks and Apache Hadoop [89] can be considered as the most important one. Hadoop provides both storage facilities, such as Hadoop Distributed File System (HDFS), spreading multiple copies of the data chunks into different machines, and parallel processing capabilities based on the MapReduce paradigm. In order to overcome some of the drawbacks Hadoop has, the Apache Spark[1] framework

---

[1] http://spark.apache.org/

may be used, especially when handling iterative jobs. Spark is a framework for cluster computing and allows for reusing a working set of data across multiple parallel operations. Compared with Hadoop it provides performance up to 100 times faster for certain application, by allowing data to be queried repeatedly [11]. Anyway, Spark cannot be used standalone but it requires an external cluster manager and a third-party distributed storage system, as it is the case with HDFS. Distributed data may be easily handled by MongoDB [2], HyperDex[3], DocumentDB[4] and the like, which are document-based database systems storing data in a JSON-like fashion. Also NoSQL databases, that exploit the key-value paradigm for distributed fast queries, may be successfully employed for managing social big data. Finally, in this context, graph-based DBs, which are capable of mirroring the social relationships amongst users, may assume a very important role. Among them, we recall Neo4J[5], Virtuoso[6], Stardog[7], etc.

Should elicitation last for at least some hours, Hadoop, Spark and NoSQL databases are surely up to the task of mining useful knowledge from social big data. On the other hand, when dealing with data streams that are quickly produced, change very quickly and need a real-time analysis, other technologies, such as Apache Storm[8] or Apache Samza[9] may be more suitable. Storm is a distributed real-time computation system useful for quick analysis of big streams of data. It is based on a master-slave approach and is composed by both a complex event processor and a distributed computation framework. Samza is a framework processing stream messages as they arrive, one at a time. Streams are divided into partitions that are in reality an ordered sequence of read-only messages.

Anyway, the aforementioned technologies must be used together with a combination of automatic machine learning, natural language elaboration, network analysis and statistics to extract interesting knowledge from social big data, the so-called *social media analytics*. As a matter of fact, text analysis algorithms, sentiment analysis and classification models may be used to detect traffic events, opinion trends and customer attitudes from streams of Status Update Messages (SUMs), comments and so on [25,44,55,57,94]. In this context, Natural Language Processing (NLP) plays a major role [19], given the fact that SUMs are typically made of text. Indeed, NLP approaches, usually based on artificial intelligence paradigms, helps machines to read and understand text by simulating the human reasoning. As a consequence careful usage of the different algorithmic steps, such as tokenization, stop-word filter-

---

[2] http://www.mongodb.com/

[3] http://hyperdex.org/

[4] http://azure.microsoft.com/en-us/services/documentdb/

[5] http://neo4j.com/product/

[6] http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSLicense

[7] http://stardog.com/

[8] http://storm.apache.org/

[9] http://samza.apache.org/

ing, stemming, stemming filtering, information extraction and so on, should be employed.

In the field of machine learning a breakthrough technology, more and more employed in the big data framework, is surely *deep learning* [20]. This emerging paradigm consists of different layers of nonlinear projections which help to learn effective representations of increasing levels of abstraction through backpropagation. Modern powerful hardware, e.g., graphics process units (GPUs), are well suited for managing millions of free parameters and parallel computations of complex architectures [72], such as Deep Belief Networks and Convolutional Neural Networks, anyway better deep learning architectures are to be devised for handling social big data in an efficient manner.

The Apache Mahout framework[10] and SAMOA[11] are examples of tools offering specific data mining algorithms for static and streaming data respectively. Apache Flink[12], which includes a machine learning library, allows handling both static big data and big data streams, and Spark has its own module for handling data streaming and a library of machine learning algorithms, namely MLib[13]. To the best of our knowledge, H_2O[14] is the most recent open source framework that provides a parallel processing engine, analytics, math, and machine learning libraries, along with data preprocessing and evaluation tools.

3.2 An operative methodology for social big data in marketing

In order to carry out social sensing and big data analytics in a marketing-effective way, we need a model taking into consideration human interactions in OSN [3, 85] and the data sources (e.g., the SUMs) of this scenario, i.e., people's posts and interplays. The major issue here concerns the correct interpretation of the huge and dynamic amount of data sources in an automatic but still correct way, exploiting the techniques described in the previous subsection. In Fig. 2 we show a cyclic operative methodology for the exploitation of social big data, integrating marketing strategies and social big data techniques. The model follows a typical approach of social media marketing and combines some interesting ideas found in the cyclic innovation model of [14]. In particular, in the shown model, big data technologies mirror Berkhout's technological research, whereas the final result reports resemble the generic product creation. Indeed, the final products are useful insights used as a feedback, phenomenon of many typical social media models, to adapt future marketing strategies.

---

[10] http://mahout.apache.org/
[11] http://samoa.incubator.apache.org/
[12] Apache Flink (2016), http://flink.apache.org
[13] Spark Mlib (2016), http://spark.apache.org/mlib
[14] h_2O (2016), http://www.h2o.ai
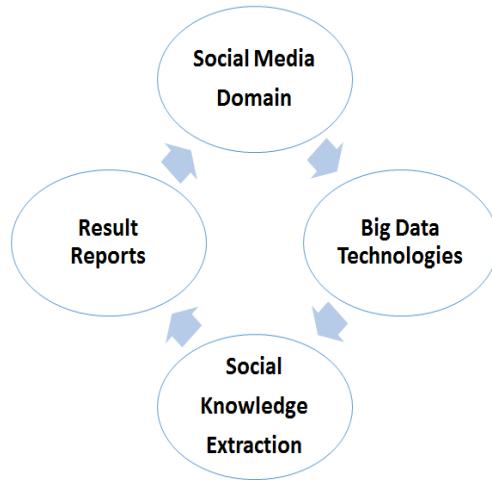
Fig. 2: A cyclic process for social big data applied to marketing.

Therefore, the results obtained at the end of the process are not set in stone and permanent, but can be helpful to consolidate and improve the strategic domain, the deployed technologies and the outcomes themselves, in future cycles. As a matter of fact, social big data analysis, and the following support to decisions, results to be more effective if accomplished continuously and adaptively, in order to monitor some strategic aspects or periodically verify the relevant changes after some activities and decisions have been enacted.

The aforementioned operative methodology encompasses four main phases, which are the following:

- definition of the strategic social media domain,
- selection of the most effective big data technologies,
- extraction and interpretation of knowledge,
- elaboration of the result reports.

The first step here is to identify the specific context, from which we want to mine useful information, e.g., the social media and social networks we intend to analyze and their most important *conversations*. These, in turn, depend on the chosen topics, contexts, markets and stackholders. At this stage, appropriate filters on the queries to the social media and social networks have to be defined. The relevant filters may regard geographic positions, the language of the conversations, temporal intervals, etc. Some specific keywords, regarding a particular company and its brands, products and competitors, may be chosen for defining appropriate queries on social media/networks.

As examined in the previous subsection, a number of technologies may be employed, for data storage, management and information mining. Furthermore, a

huge number of ready-to-use social media monitoring on-line services, such as Radian6[15], Atlas.ti[16] and T-LAB[17], are currently available for social big data analytics. Whenever we want to carry out a social media content analysis, first we have to accurately select the technologies and/or the ready-to-use monitoring services we intend to employ. Then, the social media analysis system must be setup, accordingly with the identified strategic social domain.

(a) Words cloud.   (b) Topic trend, courtesy of [93].
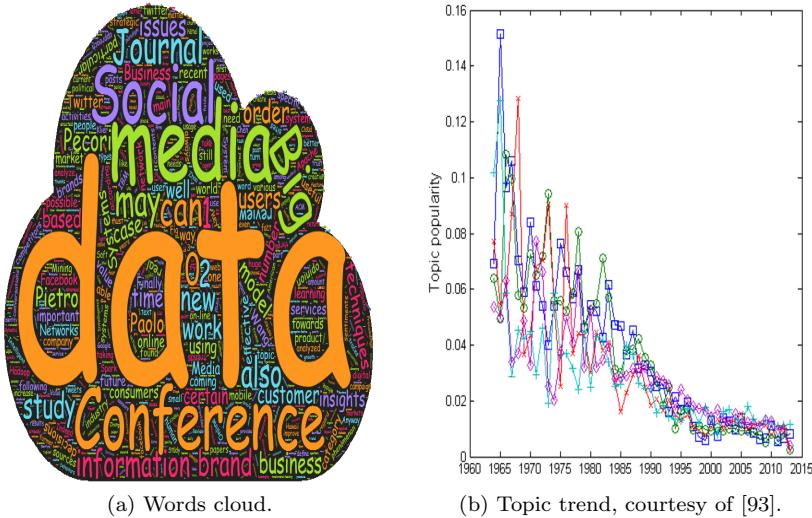
Fig. 3: Examples of types of reports.

As stated before, the stage of knowledge extraction and interpretation is strictly related to machine learning, especially to social sensing and sentiment analysis. Such paradigms, being based on passions and emotions rather than on prices and costs alone, allow to get and interpret all the information needed to build up strategic decisions. At this stage, the results should also be evaluated by a team of *content analysts* to further assess the influence of posts, the compliance of the induced sentiments to the original conveyed message and the evaluation of different types of user [86]. However, the work of the human analysts needs to be supported by result reports in order to effectively support decisions. According to the different technologies and organizational needs, various types of report can be exploited: word clouds, topic analysis and trend, influence viewer, river of news and share of conversation [21]. The *word cloud* is a set of frequent words within a given strategic domain: the bigger the word, the greater the number of times the word has been mined from

---

[15] http://www.marketingcloud.com/products/social-media-marketing/radian6/

[16] http://atlasti.com/

[17] http://tlab.it/en/presentation.php

social media/networks. The *topic analysis* concerns infographics representing, for each main topic, the distribution of the conversations where it is present. Such an action can be performed according to different parameters such as competitors, specific channels, languages, geographic provenance, sentiments, and the like. On the other hand, the *topic trend* refers to the temporal tendency of any given conversation with respect to a specific theme and/or a certain parameter. The *influence viewer* allows to detect the most influencing channels and users for certain topics. The *river of news* is simply a detailed list of conversations per single topic, while the *share of conversation* is a percentage highlighting the popularity of a certain theme in the reference domain. Two examples of the aforementioned reports are depicted in Fig. 3.

4 A Review of Social Big Data Analytics in Support of Marketing Decisions

In this section, we intend to carry out a brief analysis and review of the current literature to frame various social big data use cases within four main options when effective marketing decisions are concerned. We aim to show the best ways to emphasize growth opportunities, while highlighting the main unresolved issues in the area. We employed a systematic approach, similar to science mapping analysis [22], in order to identify the most relevant academic and peer-reviewed articles for the literature review process. First, we selected some academic databases, including IEEEXplore, Web of Science, Business Source Premier, Science Direct, Emerald, and Wiley Online Library. Then, we searched these databases using specific keywords, such as: "social media analytics", "big data mining" and "social media marketing". Finally, we selected those contributions concerning social big data analytics for marketing strategies, considering only primary studies published in English and contained in journals, conference proceedings, books and white papers. Each candidate study among such works went through the following three selection stages: 1) assessment of the title, the whole work was discarded, if its title did not prove a well-grounded relationship to social big data analysis and marketing; 2) reading of the abstract, excluding those works not touching the core topic of the research; 3) retrieval of the study and reading of its introduction and conclusions, discarding those contributions being too similar to other more relevant studies by the same authors. Moreover, in order to have a snapshot on the most recent trends, we filtered the surviving papers per year considering only 2014, 2015 and 2016. Following this phase, 52 studies remained and we proceeded to cluster those marketing actions, being present in the works we analyzed and fostered by social big data, as follows:

1. Integrating the traditional market researches and strategies allowing to highlight trends and to define new scenarios, analyze competitors and innovate products and services.

2. Developing advertisement campaigns as well as communication and pro-motional activities, both on-line and off-line.

3. Analyzing the perception and the reputation management as regards the brand, the products and the services offered by a certain firm.

4. Managing customer relationships, also known as *social customer relation-ship management*.

These categories have been chosen in order to frame separately, as much as possible, the contributions we considered, in order to get from them different open issues and research directions that will be discussed in Subsec. 4.5.

4.1 Integration of the traditional market researches and strategies

In this category, we can classify the work in [30], advocating that corporations must properly manage and adapt their physical, human and organizational resources to the big data paradigm. By 'physical means' we intend the in-volved software and platforms, the human capitals are those scientists and practitioners who know how to catch useful insights and the organizational resources specifically refer to internal processes and procedures to turn the extracted information and insights into practical daily activities. These re-sources should be carefully re-engineered, enhancing the flow of information and continuously training the personnel on the matter; this is witnessed also in [34] and in [17]. In the former case, businesses are encouraged to adopt social big data techniques, as well as mobile sensors, and incorporate their use in their economic models and strategies, as they may affect the longevity and the success of the business itself; in the latter case the importance of social media analysis is highlighted regarding the development of South-East Euro-pean companies.
The contribution in [42] shows how social big data analytics foster dynamic ca-pabilities, in order to find those hitherto unmet needs and predict consumers' behavior, as well as adaptive skills, in order to continuously measure key per-formances and maintain their temporary competitive advantage towards their competitors. In such a way, the value coming from big data could be deployed for both their incremental and radical innovations: the former may enhance both the current and the already tested marketing strategies, whereas the lat-ter allows to define new techniques, such as the new anticipatory shipping strategy employed by Amazon [6]. Another interesting example of a successful integrated marketing campaign can be drawn from [9], a work describing the social media marketing strategy of a bank. In this contribution the insights and actions coming from social big data are effectively embedded into existent business processes and legacy decision-making routines of marketing managers and all business analysts. We can also enshrine in this section the studies in [29] and [47]. In the latter work the authors try to study the engrossment, on the part of a generic business, of the knowledge and the value coming from social

Table 1: Summary of papers employing social big data as integration of traditional marketing.

Title	Authors	Source	Year	Ref.
Big data consumer analytics and the transformation of marketing	S. Erevelles, N. Fukawa, L. Swayne	Journal of Business Research	2015	[30]
Integration of mobile, big data, sensors, and social media: Impact on daily life and business	M. Gastaldi	IST-Africa Conference Proceedings	2014	[34]
The social marketing as prerequisite for the competitiveness of South-East European companies	I. Bubanja	International Convention on Information and Communication Technology, Electronics and Micro-electronics (MIPRO)	2016	[17]
Automated planning of process models: Design of a novel approach to construct exclusive choices	H. Bernd, M. Klier and S. Zimmermann	Decision Support Systems	2015	[42]
Amazon and anticipatory shipping: A dubious patent?	S. Banker	forbes.com	2014	[6]
Social Media Analytics and Business Value: A Theoretical Framework and Case Study	N. Bekmamedova and G. Shanks	Hawaii International Conference on System Sciences	2014	[9]
Managing a Big Data project: The case of Ramco Cements Limited	D. Dutta and I. Bose	International Journal of Production Economics	2015	[29]

Research on big data in business model innovation based on GA-BP model	W. Jiang and H. Chai	IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)	2016	[47]
Developing and Implementing A Social Media Program While Optimizing Return On Investment - An MBA Program Case Study	D. M. Gilfoil, S. M. Aukers and C. G. Jobs	American Journal of Business Education	2014	[36]
Introduction: Social media marketing and luxury brands	M. Phan and S.-Y. Park	Journal of Global Fashion Marketing	2014	[70]
Digital traces for business intelligence: A case study of mobile telecoms service brands in Greece	G. Milolidakis, D. Akoumianakis and C. Kimble	Journal of Enterprise Information Management	2014	[64]
The marketing effectiveness of social media in the hotel industry: a comparison of facebook and twitter	Xi Y. Leung, B. Bai and K. Stahura	Journal of Hospitality and Tourism Research	2015	[54]
What can big data and text analytics tell us about hotel guest experience and satisfaction?	Z. Xiang, Z. Schwartz, J. Gerdes and M. Uysal	International Journal of Hospitality Management	2015	[92]
Empirical Study of Using Big Data for Business Process Improvement at Private Manufacturing Firm in Cloud Computing	Z. Wang and H. Zhao	IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)	2016	[88]

big data analytics, by using genetic algorithms and back propagation neural networks, in order to fully deploy and enhance a business model made up by nine building blocks. The former highlights the complexity of a social big data project and the need for a change of mindset in both the marketing heads and the employees of any given firm, as the applying of social big data analytics would involve various phases of the production cycle, be they the strategic groundwork identification, data mining strategies and their own implementations. In this perspective, in [36] an important question is highlighted, i.e., how many resources should be re-allocated from traditional marketing tools towards new social media marketing campaign. This is expressed also in [70] where social big data are assessed in the framework of the luxury industry, and in [64] where the adoption of social big data is referred to mobile brands and where their influence, both on the internal and external organization of a company, is pointed out. The issue of firm organization is discussed also in [54], where an in-depth study of social big data, applied to the tourism industry and to its on-line marketing strategy, is presented. Another viewpoint on the matter is provided by the authors of [92], who investigate, in a qualitative and quantitative way, the implementation in the hospitality industry of the insights arising from social big data. This work uses both the data coming from social media, like comments and reviews, and tracing of customer behaviors, such as on-line web visits and off-line attitudes. Another example may be drawn from [88] where the assimilation of social big data into a Chinese manufacturing firm is analyzed through a two-layer framework.

4.2 Fostering of commercials and promotion activities

Some significant examples in this category may be those of Amazon, eBay and other similar web sites, together with their innovative selling platforms and recommender systems [11]. Some examples of these algorithms may be found also in Google Adwords [35] or IBM Business Analytics for Marketing [74], designed to help marketing leaders taking "data-driven decisions". Moreover, the work in [84] regards suggestions for ad-hoc services in order to create more and more highly-targeted advertisements and marketing campaigns, with the aim of both building business-to-customer relationships and reaching niche markets as well. This is reaffirmed also in [83], where a framework is built in order to suggest to companies the aspects and weaknesses to improve through personalized or group advertisements.
Considering social networks, even though Facebook advertisements may be perceived as unfruitful for purchasing or even annoying by most users, it is proven that they increase the returns on investments and the profits of web-based firms, through an increase in the number of visitors [39, 80]. Another recent work considering Facebook to foster marketing campaign is presented in [37], where influencing key users are detected through fuzzy C-Means clustering applied to those social big data coming from users' posts.

Important issues and challenges in this category concern the modeling of the dynamics of viral marketing on social networks, or the in-depth analysis of those feelings and passion-driven information towards certain brands [40]. In this perspective, the contribution in [16] sheds light on the possibilities of big data together with viral mobile messages for geosensing-empowered direct marketing services. This work proposes to employ a platform allowing to take advantage of the localization-based information in real time from mobile users and GIS (Geographic Information Systems) tracking technologies, in order to identify virtual market zones. Users entering a certain geofence automatically trigger personalized promotion messages. A similar work evaluates the popularity of a product, namely iPhone 6, in terms of geo-spatial coordinates and gender of the user providing tweets [43]. This leads to assess correctly short period trends and consequently to adjust eventual advertisement campaigns, according to the advantages or drawbacks of the considered product.

An innovative usage of social big data in this category is presented in [7], where the authors analyze different techniques, with different granularity degrees, for selecting appealing allies in a partner-marketing campaign on the basis of geographical, purchasing and brand information.

Specific usages of social big data in this group may also concern politics marketing, for example for driving on-line donations and voters mobilizations, policy discussions and campaign advertising, as well as transparency and participation fostering [84].

Another particular usage of social big data in this class may refer to LinkedIn [2]. Firms, through the usage of this social network, are able to observe both their customers and their potential candidates understanding what moves jobless people from the working point of view, i.e., they are able to create a general identikit of the typical millenial. The main potentiality, in this new perspective, is, however, the chance to perform an evolved recruitment.

An innovative deployment of social big data, in this cluster, can also be found in [26], where young sport talents are early identified for the sporting goods industry through social media analysis of their reputation and mentions on social networks. This task is achieved also thanks to instruments such as the aforementioned word cloud and buzz graph, as well as natural language processing tools, but it still lacks something, as the right sources of data may convey different information during time and the identified talents may quickly turn into common players.

Finally, an effective deployment of social media in this subdivision is analyzed in [41] as regards the food industry. The work presents social media as giant word-of-mouths, able to catalyze and distribute social big data virally, to establish competitive benchmarks and to monitor performances as well as possible weaknesses. The virality concept is studied also in [49] where the authors test a conceptual framework to predict pass-on behavior for online contents, and in [48], wherein a pattern based diffusion process is studied for historical retweeting behaviors and particular users are detected for fostering certain marketing campaigns.

Table 2: Summary of papers employing social big data to foster commercials.

Title	Authors	Source	Year	Ref.
Social big data: Recent achievements and new challenges	G. Bello-Orgaz, J. J. Jung and D. Camacho	Information Fusion	2016	[11]
Advanced Google AdWords	B. Geddes		2014	[35]
The SIMD accelerator for business analytics on the IBM z13	E.M. Schwarz, R.B. Krishnamurthy, C.J. Parris, J.D. Bradbury, I.M. Nnebe and M. Gschwind	IBM Journal of Research and Development	2015	[74]
How big data can make big impact: Findings from a systematic review and a longitudinal case study	S. F. Wamba, S. Akter, A. Edwards, G. Chopin and D. Gnanzou	International Journal of Production Economics	2015	[84]
TipMe: Personalized advertising and aspect-based opinion mining for users and businesses	I. Varlamis, M. Eirinaki and D. Proios	IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)	2015	[83]
Social networks, personalized advertising, and privacy controls	C. E. Tucker	Journal of Marketing Research	2014	[80]
Is bigger always better? Potential biases of big data derived from social network sites	E. Hargittai	The ANNALS of the American Academy of Political and Social Science	2015	[39]

A clustering technique to rise up the marketing tactics by looking out the key users taking Facebook as a case study	K. C. Gull, A. B. Angadi, C. G. Seema and S. G. Kanakaraddi	IEEE International Advance Computing Conference (IACC)	2014	[37]
Brands, Friends, and Viral Advertising: A Social Exchange Perspective on the Ad Referral Processes	J. L. Hayes, K. King, Karen and A. Ramirez	Journal of Interactive Marketing	2016	[40]
Viral geofencing: An exploration of emerging big-data driven direct digital marketing services	R. L. Brown, and R. R. Harmon	Portland International Conference on Management of Engineering Technology	2014	[16]
Investigation of localized sentiment for a given product by analyzing tweets	S. A. A. Hridoy, M. T. Ekram, M. S. Islam, F. Ahmed, and R. M. Rahman	IEEE/ACIS International Conference on Computer and Information Science	2015	[43]
Partner-marketing using geo-social media data for smarter commerce	J. Bao, A. Deshpande, S. McFaddin and C. Narayanaswami	IBM Journal of Research and Development	2014	[7]
Budget pacing for targeted online advertisements at LinkedIn	D. Agarwal, S. Ghosh, K. Wei, and S. You	ACM SIGKDD international conference on Knowledge discovery and data mining	2014	[2]
Identifying Sports Talents by Social Media Mining as a Marketing Instrument	P. Davcheva	Annual SRII Global Conference	2014	[26]

Enhancing social media competitiveness of small businesses: insights from small pizzerias	W. He, F. Wang, and S. Zha	New Review of Hypermedia and Multimedia	2014	[41]
The success of viral ads: Social and attitudinal predictors of consumer pass-on behavior on social network sites	P. Ketelaar et al.	Journal of Business Research	2016	[49]
Marketing Campaigns in Twitter Using a Pattern Based Diffusion Policy	E. Kafeza, C. Makris and P. Vikatos	IEEE International Congress on Big Data (BigData Congress)	2016	[48]

4.3 Analyzing perception and reputation

Concerning the reputation of any given brand, the work in [73] shows how Twitter can be an important way of spreading brands and their slogans in a sort of electronic word-of-mouth. A relevant percentage of tweets concerns some brands and a subset of these contains also sentiments towards such trademarks. These are extremely important pieces of information for corporations, whether they regard themselves or their competitors, especially because of the real-timeliness of the knowledge able to generate potential competitive advantages. The work in [58] focuses, in particular, on the time trends and frequency of tweets, considering the variation in the users' retweeting behavior according to the temporal, social and topic contexts as a whole. The authors, even if not applying any big data technology, succeed into proposing an effective predictive model for users' retweets as well as a sort of time-aware recommendation method, able to suggest to companies which information to publish during a certain time interval.

The contribution in [54] highlights that authenticity, genuineness, as well as transparency, are fundamental to appear attractive towards both new and past customers. Moreover, teamwork and passion create a more involving environment and context that may help in reducing consumers churn.

A particular case of brand reputation is that of politics and government trust.

In [4] a real time assessment of tweets, through sentiment analysis and various classification algorithms, has been successfully deployed in two scenarios: the US 2012 Presidential Elections and the Karnataka 2013 Assembly Elections. The effectiveness in monitoring public opinion and forecasting political results emerges also in [75], where Twitter has been exploited to mine dynamic social trends and to highlight communications among people supporting the same political candidates, and in [18], where the authors try to find a relation between big data analytics coming from Twitter and the citizens' trust towards politicians and government in the aftermath of the 2014 Brazilian World Cup. An interesting analysis about electoral big data and their distributions can be found also in [87], where historical elections data and potential policy gains are investigated through fuzzy methods in order to assess policy confirmations and general consensus.

The work in [96] exploits sentiment analysis and users' perceptions as well; in this case the objective is to create an intelligent recommending system that is emotion-aware and able to rate revisions on the basis of particular emotional offsets. A further work based on sentiment analysis in this category is the one in [59], where the authors try to improve location-based recommending systems through sentiments and opinions of users in some microblogs about certain point of interests. Considering the sentiment attributes of locations can result into an improvement of the performance for travel agencies and websites, through more humanization and closeness to what consumers really need and feel.

A different specific application of brand reputation through social big data can be found again in LinkedIn [2]. This social network is able to provide useful social big data both for companies and for those people looking for a job position, because enterprises have to promote their business brand whereas unemployed people want to advertise their personal brand.

Further examples in this category can be found in [9], [50] and in [66]. In the first of these works the authors analyze the improvement of brand reputation in the financial sector, specifically a bank tried to change the impression consumers have towards itself exploiting new social media channels. Also the second work considers the enhancing of web pages of some banks in the perspective of increasing or maintaining a brand loyalty; this is accomplished through an eye tracker analysis of Facebook pages and the usage of heat maps. In the third aforementioned work, the researchers propose an assessment of trends and reputations of TV programs on social channels, correlating audience ratings during certain periods of time with the amount of social big data generated over particular social networks concerning the same program or TV series. The convergence between traditional TV indicators and social big data is studied also in [69], where the authors try to get much deeper real-time understanding of what viewers think about shows and the brands that advertise on them, through a concept-level integration framework exploiting conceptual abstractions and ontological domain knowledge.

In [77] the authors present a product improvement framework based on social media analytics, focusing especially on Twitter as a source of opinions and

sentiments of users in order to improve a company's product, namely Apple iPhone.

Finally, in this category we can also mention the usage of social big data to find opportune influencers for a certain business or product [52]. For example, fashion influencers and fashion bloggers are an emerging phenomenon of the last years: they are capable of driving trends and purchases only through opportune photos and proper comments about some brands. Therefore discovering correct people for increasing the reputation of a certain product is going to become one of the main usage of social big data analytics.

4.4 Managing customer relationship

The study in [71] presents a recent project, with the aim of bringing together enterprises and consumers. The main objective is to increase the return of investments of corporations through accurate social media marketing campaigns. This is achieved through an involvement strategy, targeted to consumers, taking advantage of gaming techniques. Thereafter the user and their information are sold as the main sale product. The proposed platform aims to be a concrete framework for social media marketing, capable of positioning firms and geofencing markets, but it deploys a holistic approach and it undergoes some important and not trivial market barriers such as penetrating the global audience.

An innovative and upstream idea for this category can be found in [12] as well. Here, the authors envision consumers being paid with incentives or real money for their participation in online communities, in leaving comments, and so on. These ideas are indeed already used by some web services providing cashbacks, both in actual money and through coupons and vouchers, for on-line purchases passing through their web pages or for participating into surveys or pay-per-click campaigns. Some examples of such sites are Beruby.com[18], i-say.com[19], Nielsen Online Rewards[20], and the like. In [41], it is highlighted that it is possible to foster customer care, customer support and retention, as well as development and innovation, driven by the consumers themselves, through praises and complaints. Another important aspect regards the creation of a friendly and happy atmosphere, by means of instruments such as greetings, promotions, deals, contests, games and so on, that boosts the attraction of new customers, their loyalty to the company brand and the mitigation of certain levels of criticality [33]. Thanks to their social traces, customers can be in the middle of the marketing strategy and become more powerful and participative in the company marketing decisions. This is mainly due to the public and transparent features of social big data compared with the private opportunities for complaining and praising that were available in the past.

---

[18] http://us.beruby.com/

[19] http://i-say.com/

[20] https://nielsenonlinerewards.com/

Table 3: Summary of papers employing social big data used to analyze perception and reputation of brands.

Title	Authors	Source	Year	Ref.
Entifying your brand among Twitter-using millennials	H. Sashittal, M. Hodis and R. Sriramachandramurthy	Business Horizons	2015	[73]
Discovering Temporal Retweeting Patterns for Social Media Marketing Campaigns	G. Liu, Y. Fu, T. Xu, H. Xiong and G. Chen	IEEE International Conference on Data Mining	2014	[58]
The marketing effectiveness of social media in the hotel industry: a comparison of facebook and twitter	Xi Y. Leung, B. Bai and K. Stahura	Journal of Hospitality and Tourism Research	2015	[54]
Influence factor based opinion mining of Twitter data using supervised learning	M. Anjaria and R.M.R. Guddeti	International Conference on Communication Systems and Networks	2014	[4]
Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter	M. Song, M. C. Kim and Y. K. Jeong	IEEE Intelligent Systems	2014	[75]
Big Data Analyses for Collective Opinion Elicitation in Social Networks	Y. Wang and V. J. Wiebe	IEEE International Conference on Trust, Security and Privacy in Computing and Communications	2014	[87]

Mixed-initiative social media analytics at the World Bank: Observations of citizen sentiment in Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest	N. A. Calderon et al.	IEEE International Conference on Big Data	2015	[18]
GroRec: A Group-centric Intelligent Recommender System Integrating Social, Mobile and Big Data Technologies	Y. Zhang	IEEE Transactions on Services Computing	2016	[96]
Schedule a Rich Sentimental Travel via Sentimental POI Mining and Recommendation	P. Lou, G. Zhao, X. Qian, H. Wang and X. Hou	IEEE Second International Conference on Multimedia Big Data (BigMM)	2016	[59]
Budget pacing for targeted online advertisements at LinkedIn	D. Agarwal, S. Ghosh, K. Wei, and S. You	ACM SIGKDD international conference on Knowledge discovery and data mining	2014	[2]
Social Media Analytics and Business Value: A Theoretical Framework and Case Study	N. Bekmamedova and G. Shanks	Hawaii International Conference on System Sciences	2014	[9]

A comparative study of user experience in online social media branding web pages using eye tracker	N. Kumar, V. Maheshwari and J. Kumar	International Conference on Advances in Human Machine Interaction (HMI)	2016	[50]
Is Twitter psychic? Social media analytics and television ratings	C. Oh et al.	International Conference on Computing Technology and Information Management	2015	[66]
Leveraging Cross-Domain Social Media Analytics to Understand TV Topics Popularity	R. G. Pensa, M. L. Sapino, C. Schifanella and L. Vignaroli	IEEE Computational Intelligence Magazine	2016	[69]
Social Media Analytics Based Product Improvement Framework	C. J. Su and Y. A. Chen	International Symposium on Computer, Consumer and Control (IS3C)	2016	[77]
How big data helps to find influencers on Twitter. An empirical study	E. Lauherta-Otero, and R. Cordero-Gutierrez	AEMARK Conference)	2016	[52]

The central role of customers' social big data is highlighted also in [81], where an accurate analysis of users' engagement is carried out on some Facebook pages, considering the number of likes, of posts, of shares and of comments, in the latter case both as voluntary actions and as a response to the moderator's ones. In the work it is found that the interactivity and the quality of posts may be much more successful than the bare number of posts and of fans.

Finally, the work in [5] suggests to fruitful employ social big data analytics to identify strategies for retaining customers before they decide to leave a company for a competitor. This fits perfectly the logic of the *customer lifetime value*, defined as the present value of all existing and future cash flows generated from a customer [95]. As a matter of fact, the cost of gaining a new client

is significantly higher than that of avoiding his/her churn in favor of competitors. In this perspective, the analysis of social big data may help corporations to be proactive and providing offers and discounted prices for long term dissatisfied customers. However, identifying the real cause of customers' churn, such as inefficiencies or lost time, is not always easy. Indeed, understanding real motivations is a challenging task and may not always yield users' true intentions, thus wrong predictions may still take place and strategies to avoid them should be enacted. This difficulty is evident also in the study in [23], where the links between brand equity, consumers' engagement and purchase intentions are thoroughly analyzed. The findings of this work are mainly the importance of creating a community connected to the brand and the central role of brand equity that can both be influenced by engaging social media contents and foster a further involvement into social media activities. The same authors in [24] analyze further the relationship between social big data content types, the benefits perceived by the consumers involved and their connection with loyalty and purchase intentions. The study, which considers both a luxury and a non-luxury French cosmetics brand, reveals that only social and hedonic benefits, deriving from dialog and community contents respectively, can be useful predictors of future consumers' expenditures, whereas informational benefits, coming from news content, are not so influencing.

4.5 Discussion and future research

The following figures can be drawn from the recent literature we reviewed: 14 papers deal with the integration of social big data into traditional marketing techniques, 17 address the fostering of promotions through social big data, 16 try to exploit social big data in the perspective of brand reputation and finally 9 concern the customer relationship. Only 4 papers of 52 can be classified in more than one category, thus the considered classification is quite effective in differentiating the various points of view of marketing and social big data. As it can be inferred from Fig. 4, the analyzed research production on the topic is slightly decreasing from the 20 works of 2014 to the 17 of 2016, but the maturity of the studies is keeping constant as one can see from the number of journal publications, numbering about 7 during every year we considered, while the number of conference contributions experienced a slight decrease during the same period. The scientific research in the field has reached a certain maturity overall, as the number of journal contributions is almost the half of the analyzed papers (22), anyway some interesting conference papers are still being published, as they encompass ongoing developments by analyzing in depth some subtle details, remained uncovered till now, such as the most involving contents, on the part of the consumers, of social big data or the usage of this pervasive information to recruit new effective testimonials.

Table 4: Summary of papers employing social big data used to foster customer relationship.

Title	Authors	Source	Year	Ref.
A big data aggregation, analysis and exploitation integrated platform for increasing social management intelligence	G. Sand at al.	IEEE International Conference on Big Data	2014	[71]
A Review of Information Systems Research on Online Social Networks	K. Berger, J. Klier, M. Klier and F. Probst	Communications of the Association for Information Systems	2014	[12]
Enhancing social media competitiveness of small businesses: insights from small pizzerias	W. He, F. Wang, and S. Zha	New Review of Hypermedia and Multimedia	2014	[41]
E-reputation: A case study of organic cosmetics in social media	F. Fourati-Jamoussi	International Conference on Information Systems and Economic Intelligence	2015	[33]
Digital brand management - A study on the factors affecting customers' engagement in Facebook pages	V. M. Vadivu and M. Neelamalar	International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)	2015	[81]
Analytics: Key to Go from Generating Big Data to Deriving Business Value	D. Arora and P. Malik	IEEE International Conference on Big Data Computing Service and Applications	2015	[5]
Influence of customer engagement with company social networks on stickiness: Mediating effect of customer value creation	Z. Mingli, G. Lingyun, H. Mu and L. Wenhua	International Journal of Information Management	2016	[95]

Do Facebook Likes Lead to Shares or Sales? Exploring the Empirical Links between Social Media Content, Brand Equity, Purchase Intention, and Engagement	C. K. Coursaris, W. v. Osch and B. A. Balogh	Hawaii International Conference on System Sciences (HICSS)	2016	[23]
Disentangling the drivers of brand benefits: Evidence from social media marketing of two cosmetics companies	C. K. Coursaris and W. Van Osch	International Conference on Telecommunications and Multimedia (TEMU)	2016	[24]
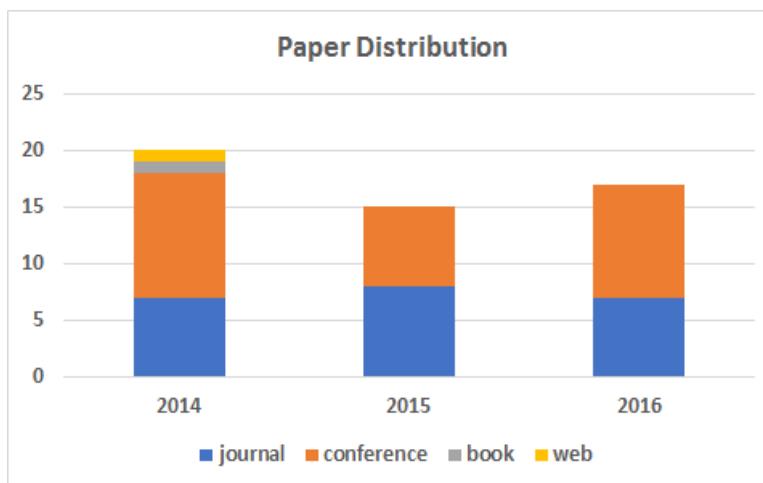
Fig. 4: Distribution of the analyzed papers.

In general, the four considered categories, used to classify the recent literature, suggest different findings and open issues, summarized in the following.

As regards the integration of legacy marketing techniques there is the need for a change in company cultures towards the more dynamic and adaptive stream of social big data, in which corporations should delve and imbibe themselves. This should take place, in the best possible way, in a collaborative and distributed way across different platforms, for example through the usage of system of systems service arhcitectures as envisoned in [90].
Furthermore, social big data must change the marketing strategies without turning upside-down the whole organization; what appears from the survey is the importance to focus on a few social marketing channels, as the point is not to please everybody, but rather the most fruitful consumers.
Finally, the reliability of the results mainly depends on the maturity of the local market and clients: what is interesting, and pointed out also by [79], is that social big data sources and contents may vary between western and eastern markets and different prerequisites must be considered in order to correctly deploy their insights.

Concerning the fostering of advertisement what can be gleaned from this survey is that variables such as the penetration degree, the transmission rate and the emotive depth of marketing campaigns should be carefully considered in a possibly unified framework. Moreover, the reasons for the happening of spikes of emotions during certain time intervals and the analysis of different on-line social networks, other than the traditional Facebook and Twitter, are some possible research gaps in this field that the authors encourage to investigate further.

The suggestions coming from the third category we considered concern the fact that the proposed social big data analysis techniques could be deployed to investigate public opinion about hot social and political issues as well, in order to drive politicians and companies towards better decisions for the whole society, which in turn could boost brand and identity reputation. This could be enacted also considering legacy TV channels as a tool complementary to modern social media networks, but this integration needs further research.

From the fourth category we can draw the conclusion that penetrating the global audience is not so easy, even with the help of social big data, and that the bond with the customers could be strengthened involving them in social activities, such as polls and questionnaires promoted by the companies themselves. Anyway, finding the real cause of customers' churn, such as inefficiencies or lost time, is not always easy: understanding the real motivations and the effective drivers of increasing purchases is a challenging task and may not always yield users' true intentions, thus wrong predictions still may take place and strategies to avoid them should be enacted.

The whole insights of the analyses we carried out throughout this paper concern the power of social big data to enable companies: i) to garner suggestions

and hints useful in defining new marketing strategies, ii) to adapt the existing strategies to a new context undertaking corrective actions, iii) to determine new or past scenarios where to intervene in terms of communications, products, services and markets. What comes to light is that it is desirable to extend further the intersections among social media, big data and data analysis techniques, maybe through a combination of different social big data dimensions, and sentiment analysis into the same marketing methods. This is motivated by the fact that mixed methods, which take advantage of contents, relations, experiences and activities, improve the confidence marketing managers have in their findings, and they can also yield valuable insights into the inadequacies, gaps, biases of traditional one-dimension procedures [8].

However, some conclusive issues to be investigated further, concerning all the categories considered, are the following: i) the doubt whether social big data activities should be outsourced rather than conducted within the company itself and in which case the return shall prove to be greater, ii) the required number of experts and scientists, with a deep knowledge of social big data analysis techniques, for a successful marketing strategy, iii) the need for a standardized assessment platform and the lack of an integrated theoretical framework for both social media analytics and Big Data technologies, iv) the necessity to take particular care with privacy and security issues as well as an effective risk assessment [78] v) the growing need for real-time and timely processing of social big data and vi) a lexicon-independent sentiment analysis [32].

5 Conclusions and future perspectives

In this paper, we surveyed social big data, the current techniques to manage them and a variety of effective marketing deployments based on such a huge amount of data. In the first part we have accurately described the numbers characterizing *Society 2.0*, the possible dimensions of social big data and the technologies currently employed to handle them. Moreover, we described an effective and operative methodology to foster marketing strategies by taking advantage of such valuable information sources that social big data represent. In the second part, we provided a systematic review of the recent literature on social big data and marketing, uncovering some findings and some open research issues.

What can be concluded is that social big data is no more a niche topic nowadays, but rather a proper phenomenon capable of radically changing the world and to promote a great variety of benefits such as the increase of operating margin, the growth in the available jobs and working positions, the growth in market figures, the escalation of on-time decisions and saved time, the boost in digitally-generated data and customer experience-driven promotions and the like.

After a broad-spectrum analysis of the main marketing possibilities of social big data in various fields and their classification, we have to conclude that we are still far from a single centralized, holistic and rigorous platform having a few buttons through which a marketer can take relevant decisions for his future market strategies. Indeed, the aims and the fields of application may be very different and, in each context, many issues and problems still have to be overcome and investigated further.

Acknowledgments

Compliance with ethical standards

Conflict of interest P. Ducange, R. Pecori, P. Mezzina declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. What happens online in 60 seconds? (2014). URL `https://clt.vtc.edu.hk/what-happens-online-in-60-seconds/`
2. Agarwal, D., Ghosh, S., Wei, K., You, S.: Budget pacing for targeted online advertisements at linkedin. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1613–1619. ACM (2014)
3. Anastasi, G., Antonelli, M., Bechini, A., Brienza, S., D'Andrea, E., De Guglielmo, D., Ducange, P., Lazzerini, B., Marcelloni, F., Segatori, A.: Urban and social sensing for sustainable mobility in smart cities. In: Sustainable Internet and ICT for Sustainability (SustainIT), 2013, pp. 1–4. IEEE (2013)
4. Anjaria, M., Guddeti, R.: Influence factor based opinion mining of twitter data using supervised learning. In: 2014 Sixth International Conference on Communication Systems and Networks, pp. 1–8 (2014)
5. Arora, D., Malik, P.: Analytics: Key to go from generating big data to deriving business value. In: 2015 IEEE First International Conference on Big Data Computing Service and Applications, pp. 446–452 (2015)
6. Banker, S.: Amazon and anticipatory shipping: A dubious patent? (2014). URL `http://www.forbes.com/sites/stevebanker/2014/01/24/amazon-and-anticipatory-shipping-a-dubious-patent`
7. Bao, J., Deshpande, A., McFaddin, S., Narayanaswami, C.: Partner-marketing using geo-social media data for smarter commerce. IBM Journal of Research and Development **58**(5/6), 6:1–6:12 (2014)
8. Behrendt, S., Richter, A., Trier, M.: Mixed methods analysis of enterprise social networks. Computer Networks **75, Part B**, 560 – 577 (2014). Special Issue on Online Social NetworksThe Connectedness, Pervasiveness and Ubiquity of Online Social Networks

9. Bekmamedova, N., Shanks, G.: Social media analytics and business value: A theoretical framework and case study. In: 2014 47th Hawaii International Conference on System Sciences, pp. 3728–3737 (2014). DOI 10.1109/HICSS.2014.464
10. Bell, D.: The Coming Of Post-industrial Society. Basic Books (2008)
11. Bello-Orgaz, G., Jung, J.J., Camacho, D.: Social big data: Recent achievements and new challenges. Information Fusion **28**, 45 – 59 (2016)
12. Berger, K., Klier, J., Klier, M., Probst, F.: A review of information systems research on online social networks. Communications of the Association for Information Systems **35**(1), 145–172 (2014)
13. Berkhout, A.J., et al.: Innovating the innovation process. International Journal of Technology Management **34** (2006)
14. Berkhout, G., Hartmann, D., Trott, P.: Connecting technological capabilities with market needs using a cyclic innovation model. R&d Management **40**(5), 474–490 (2010)
15. Bowen, G., Bowen, D.: Social media: Strategic decision making tool. Competitive Social Media Marketing Strategies p. 94 (2016)
16. Brown, R., Harmon, R.: Viral geofencing: An exploration of emerging big-data driven direct digital marketing services. In: 2014 Portland International Conference on Management of Engineering Technology, pp. 3300–3308 (2014)
17. Bubanja, I.: The social marketing as prerequisite for the competitiveness of southeast european companies. In: 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1552–1556 (2016)
18. Calderon, N.A., Fisher, B., Hemsley, J., Ceskavich, B., Jansen, G., Marciano, R., Lemieux, V.L.: Mixed-initiative social media analytics at the world bank: Observations of citizen sentiment in twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. In: 2015 IEEE International Conference on Big Data, pp. 1678–1687 (2015)
19. Cambria, E., White, B.: Jumping nlp curves: a review of natural language processing research [review article]. IEEE Computational Intelligence Magazine **9**(2), 48–57 (2014)
20. Chen, X.W., Lin, X.: Big data deep learning: Challenges and perspectives. IEEE Access **2**, 514–525 (2014). DOI 10.1109/ACCESS.2014.2325029
21. Chi, M.T., Lin, S.S., Chen, S.Y., Lin, C.H., Lee, T.Y.: Morphable word clouds for time-varying text data visualization. IEEE Transactions on Visualization and Computer Graphics **21**(12), 1415–1426 (2015). DOI 10.1109/TVCG.2015.2440241
22. Cobo, M., Lopez-Herrera, A., Herrera-Viedma, E., Herrera, F.: Science mapping software tools: Review, analysis and cooperative study among tools. Journal of the American Society for Information Science and Technology **62**(7), 1382–1402 (2011). DOI doi:10.1002/asi.2152
23. Coursaris, C.K., v. Osch, W., Balogh, B.A.: Do facebook likes lead to shares or sales? exploring the empirical links between social media content, brand equity, purchase intention, and engagement. In: 2016 49th Hawaii International Conference on System Sciences (HICSS), pp. 3546–3555 (2016)
24. Coursaris, C.K., Osch, W.V.: Disentangling the drivers of brand benefits: Evidence from social media marketing of two cosmetics companies. In: 2016 International Conference on Telecommunications and Multimedia (TEMU), pp. 1–6 (2016)
25. D'Andrea, E., Ducange, P., Lazzerini, B., Marcelloni, F.: Real-time detection of traffic from twitter stream analysis. IEEE Transactions on Intelligent Transportation Systems **16**(4), 2269–2283 (2015)
26. Davcheva, P.: Identifying sports talents by social media mining as a marketing instrument. In: 2014 Annual SRII Global Conference, pp. 223–227 (2014)
27. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. Communications of the ACM **51**(1), 107–113 (2008)
28. Ducange, P., Pecori, R., Sarti, L., Vecchio, M.: Educational Big Data Mining: How to Enhance Virtual Learning Environments, pp. 681–690. Springer International Publishing, Cham (2017)
29. Dutta, D., Bose, I.: Managing a big data project: The case of ramco cements limited. International Journal of Production Economics **165**, 293 – 306 (2015)
30. Erevelles, S., Fukawa, N., Swayne, L.: Big data consumer analytics and the transformation of marketing. Journal of Business Research **69**(2), 897 – 904 (2016)

31. Erl, T., Khattak, W., Buhler, P.: Big Data Fundamentals: Concepts, Drivers & Techniques. Prentice Hall Press (2016)
32. Fiarni, C., Maharani, H., Pratama, R.: Sentiment analysis system for indonesia online retail shop review using hierarchy naive bayes technique. In: 2016 4th International Conference on Information and Communication Technology (ICoICT), pp. 1–6 (2016)
33. Fourati-Jamoussi, F.: E-reputation: A case study of organic cosmetics in social media. In: 2015 6th International Conference on Information Systems and Economic Intelligence, pp. 125–132 (2015)
34. Gastaldi, M.: Integration of mobile, big data, sensors, and social media: Impact on daily life and business. In: IST-Africa Conference Proceedings, 2014, pp. 1–10 (2014)
35. Geddes, B.: Advanced Google AdWords. John Wiley and Sons (2014)
36. Gilfoil, D.M., Aukers, S.M., Jobs, C.G.: Developing and implementing a social media program while optimizing return on investment - an mba program case study. American Journal of Business Education **8**(1), 31–48 (2014)
37. Gull, K.C., Angadi, A.B., Seema, C.G., Kanakaraddi, S.G.: A clustering technique to rise up the marketing tactics by looking out the key users taking facebook as a case study. In: 2014 IEEE International Advance Computing Conference (IACC), pp. 579–585 (2014)
38. Hamed, A., Wu, X.: Does social media big data make the world smaller? an exploratory analysis of keyword-hashtag networks. In: 2014 IEEE International Congress on Big Data, pp. 454–461 (2014)
39. Hargittai, E.: Is bigger always better? potential biases of big data derived from social network sites. The ANNALS of the American Academy of Political and Social Science **659**(1), 63–76 (2015)
40. Hayes, J.L., King, K.W., Ramirez, A.: Brands, friends, and viral advertising: A social exchange perspective on the ad referral processes. Journal of Interactive Marketing **36**, 31–45 (2016)
41. He, W., Wang, F.K., Zha, S.: Enhancing social media competitiveness of small businesses: insights from small pizzerias. New Review of Hypermedia and Multimedia **20**(3), 225–250 (2014)
42. Heinrich, B., Klier, M., Zimmermann, S.: Automated planning of process models: Design of a novel approach to construct exclusive choices. Decision Support Systems **78**, 1 – 14 (2015)
43. Hridoy, S., Ekram, M., Islam, M., Ahmed, F., Rahman, R.: Investigation of localized sentiment for a given product by analyzing tweets. In: 2015 IEEE/ACIS 14th International Conference on Computer and Information Science, pp. 265–270 (2015)
44. Ibrahim, O.A.S., Landa-Silva, D.: Term frequency with average term occurrences for textual information retrieval. Soft Computing **20**(8), 3045–3061 (2016)
45. Jamshidi, M., Tannahill, B., Yetis, Y., Kaplan, H.: Big data analytic via soft computing paradigms. In: Frontiers of Higher Order Fuzzy Sets, pp. 229–258. Springer (2015)
46. Jha, A., Dave, M., Madan, S.: A review on the study and analysis of big data using data mining techniques. International Journal of Latest Trends inEngineering and Technology (IJLTET) **6**(3) (2016)
47. Jiang, W., Chai, H.: Research on big data in business model innovation based on ga-bp model. In: 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), pp. 174–177 (2016)
48. Kafeza, E., Makris, C., Vikatos, P.: Marketing campaigns in twitter using a pattern based diffusion policy. In: 2016 IEEE International Congress on Big Data (BigData Congress), pp. 125–132 (2016)
49. Ketelaar, P.E., Janssen, L., Vergeer, M., van Reijmersdal, E.A., Crutzen, R., vant Riet, J.: The success of viral ads: Social and attitudinal predictors of consumer pass-on behavior on social network sites. Journal of Business Research **69**(7), 2603–2613 (2016)
50. Kumar, N., Maheshwari, V., Kumar, J.: A comparative study of user experience in online social media branding web pages using eye tracker. In: 2016 International Conference on Advances in Human Machine Interaction (HMI), pp. 1–8 (2016)
51. Laney, D.: 3D data management: Controlling data volume, velocity, and variety. Tech. rep., META Group (2001)
52. Lauherta-Otero, E., Cordero-Gutierrez, R.: How big data helps to find influencers on twitter. an empirical study. In: AEMARK Conference, pp. 627–641 (2016)

53. Lesk, M.: Big data, big brother, big money. IEEE Security Privacy **11**(4), 85–89 (2013)
54. Leung, X.Y., Bai, B., Stahura, K.A.: The marketing effectiveness of social media in the hotel industry: a comparison of facebook and twitter. Journal of Hospitality and Tourism Research **39**(2), 147–169 (2015)
55. Li, J., Fong, S., Zhuang, Y., Khoury, R.: Hierarchical classification in text mining for sentiment analysis of online news. Soft Computing **20**(9), 3411–3420 (2016)
56. Lin, D., Li, S., Cao, D.: Making intelligent business decisions by mining the implicit relation from bloggers posts. Soft Computing **14**(12), 1317–1327 (2010)
57. Liu, B.: Sentiment analysis and opinion mining. Synthesis lectures on human language technologies **5**(1), 1–167 (2012)
58. Liu, G., Fu, Y., Xu, T., Xiong, H., Chen, G.: Discovering temporal retweeting patterns for social media marketing campaigns. In: 2014 IEEE International Conference on Data Mining, pp. 905–910 (2014)
59. Lou, P., Zhao, G., Qian, X., Wang, H., Hou, X.: Schedule a rich sentimental travel via sentimental poi mining and recommendation. In: 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), pp. 33–40 (2016)
60. Lu, Q., Li, Z., Zhang, W., Yang, L.T.: Autonomic deployment decision making for big data analytics applications in the cloud. Soft Computing pp. 1–12 (2015)
61. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt (2013)
62. McCafferty, D.: Surprising statistics about big data (2014). URL `http://www.baselinemag.com/analytics-big-data/slideshows/surprising-statistics-about-big-data.html`
63. Milgram, S.: The small world problem. Psychology today **2**(1), 60–67 (1967)
64. Milolidakis, G., Akoumianakis, D., Kimble, C.: Digital traces for business intelligence: A case study of mobile telecoms service brands in greece. Journal of Enterprise Information Management **27**(1), 66–98 (2014)
65. Moreno, M.V., Dufour, L., Skarmeta, A.F., Jara, A.J., Genoud, D., Ladevie, B., Bezian, J.J.: Big data: the key to energy efficiency in smart buildings. Soft Computing **20**(5), 1749–1762 (2016)
66. Oh, C., Yergeau, S., Woo, Y., Wurtsmith, B., Vaughn, S.: Is twitter psychic? social media analytics and television ratings. In: Second International Conference on Computing Technology and Information Management (ICCTIM 2015), pp. 150–155 (2015)
67. Owais, S.S., Hussein, N.S.: Extract five categories cpivw from the 9vs characteristics of the big data. International Journal of Advanced Computer Science and Applications **7**(3), 254–258 (2016)
68. Pecori, R.: S-kademlia: A trust and reputation method to mitigate a sybil attack in kademlia. Computer Networks **94, Part C**, 205–218 (2016)
69. Pensa, R.G., Sapino, M.L., Schifanella, C., Vignaroli, L.: Leveraging cross-domain social media analytics to understand tv topics popularity. IEEE Computational Intelligence Magazine **11**(3), 10–21 (2016)
70. Phan, M., Park, S.Y.: Introduction: Social media marketing and luxury brands. Journal of Global Fashion Marketing **5**(3), 195–196 (2014)
71. Sand, G., Tsitouras, L., Dimitrakopoulos, G., Chatzigiannakis, V.: A big data aggregation, analysis and exploitation integrated platform for increasing social management intelligence. In: 2014 IEEE International Conference on Big Data, pp. 40–47 (2014)
72. Sani, L., et al.: Efficient Search of Relevant Structures in Complex Systems, pp. 35–48. Springer International Publishing, Cham (2016). DOI 10.1007/978-3-319-49130-1_4
73. Sashittal, H.C., Hodis, M., Sriramachandramurthy, R.: Entifying your brand among twitter-using millennials. Business Horizons **58**(3), 325–333 (2015)
74. Schwarz, E., Krishnamurthy, R., Parris, C., Bradbury, J., Nnebe, I., Gschwind, M.: The SIMD accelerator for business analytics on the IBM z13. IBM Journal of Research and Development **59**(4/5), 2:1–2:16 (2015)
75. Song, M., Kim, M.C., Jeong, Y.K.: Analyzing the political landscape of 2012 korean presidential election in twitter. IEEE Intelligent Systems **29**(2), 18–26 (2014)
76. Song, P., Phang, C.W.: Promoting continuance through shaping members' social identity in knowledge-based versus support/advocacy virtual communities. IEEE Transactions on Engineering Management **63**(1), 16–26 (2016)

77. Su, C.J., Chen, Y.A.: Social media analytics based product improvement framework. In: 2016 International Symposium on Computer, Consumer and Control (IS3C), pp. 393–396 (2016)
78. Tanimoto, S., Iwashita, M., Seki, Y., Sato, H., Kanai, A.: Risk assessment quantification of social-media utilization in enterprise. In: 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), pp. 1–6 (2016)
79. To, W.M., Lai, L.: Data analytics in china: Trends, issues, and challenges. IT Professional **17**(4), 49–55 (2015)
80. Tucker, C.E.: Social networks, personalized advertising, and privacy controls. Journal of Marketing Research **51**(5), 546–562 (2014)
81. Vadivu, V.M., Neelamalar, M.: Digital brand management - a study on the factors affecting customers' engagement in facebook pages. In: 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 71–75 (2015)
82. del Val, E., Martínez, C., Botti, V.: Analyzing users' activity in online social networks over time through a multi-agent framework. Soft Computing pp. 1–15 (2016)
83. Varlamis, I., Eirinaki, M., Proios, D.: Tipme: Personalized advertising and aspect-based opinion mining for users and businesses. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1489–1494 (2015)
84. Wamba, S.F., Akter, S., Edwards, A., Chopin, G., Gnanzou, D.: How big data can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics **165**, 234–246 (2015)
85. Wang, D., Kaplan, L., Abdelzaher, T., Aggarwal, C.: On credibility estimation tradeoffs in assured social sensing. IEEE Journal on Selected Areas in Communications **31**(6), 1026–1037 (2013)
86. Wang, X., Luo, X.: Sentimental space based analysis of user personalized sentiments. In: 2013 Ninth International Conference on Semantics, Knowledge and Grids, pp. 151–156 (2013)
87. Wang, Y., Wiebe, V.J.: Big data analyses for collective opinion elicitation in social networks. In: 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 630–637 (2014)
88. Wang, Z., Zhao, H.: Empirical study of using big data for business process improvement at private manufacturing firm in cloud computing. In: 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud), pp. 129–135 (2016)
89. White, T.: Hadoop: The definitive guide, 4th edn. O'Reilly Media (2015)
90. Wong, R.K., Chi, C.H., Yu, Z., Zhao, Y.: A system of systems service design for social media analytics. In: 2014 IEEE International Conference on Services Computing (SCC), pp. 789–796 (2014)
91. Wu, X., Zhu, X., Wu, G.Q., Ding, W.: Data mining with big data. IEEE transactions on knowledge and data engineering **26**(1), 97–107 (2014)
92. Xiang, Z., Schwartz, Z., Gerdes, J.H., Uysal, M.: What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management **44**, 120–130 (2015)
93. Yan, E.: Research dynamics, impact, and dissemination: A topic-level analysis. JASIST **66**, 2357–2372 (2015)
94. Zeng, D., Chen, H., Lusch, R., Li, S.H.: Social media analytics and intelligence. IEEE Intelligent Systems **25**(6), 13–16 (2010)
95. Zhang, M., Guo, L., Hu, M., Liu, W.: Influence of customer engagement with company social networks on stickiness: Mediating effect of customer value creation. International Journal of Information Management (2016)
96. Zhang, Y.: Grorec: A group-centric intelligent recommender system integrating social, mobile and big data technologies. IEEE Transactions on Services Computing **PP**(99), 1–1 (2016)