Topic Modeling Analysis of Social Enterprises: Twitter Evidence

Shr-Wei Kao * and Pin Luarn

Department of Business Administration, National Taiwan University of Science and Technology, Taipei City 10607, Taiwan; luarn@mail.ntust.edu.tw
* Correspondence: welles@chtf.org.tw

Received: 21 February 2020; Accepted: 17 April 2020; Published: 22 April 2020

Abstract: Social media is a major channel used for communication by professional and social groups. The text posted on social media contains extremely rich information. To capture the development of social enterprises (SEs), this paper examines the tweets posted on Twitter and searches the hashtags on the Twitter Application Programming Interface (API) that SEs deem to be the most important. The results suggest that these tweets can be divided into three content groups (strategy, impact and business). This paper expands this into four dimensions (strategy, impact, business and people) and six indicators (social, opportunity, change, enterprise, network and team) and establishes a conceptual framework of SEs. This paper aims to enhance the understanding of the pertinent issues recently affecting SEs and extract findings that can act as a reference for follow-up studies.

Keywords: social enterprises; sustainability; topic modeling; Twitter; business model; social innovation; social entrepreneurship

1. Introduction

Social enterprises (SEs) represent a new business model that marries social value with corporate profits. They create a profit in the normal course of business and translate the profit into their own funding source. In other words, SEs are self-sufficient and sustainable entities. This paper believes that the development of SEs deserves academic attention. Battilana and Lee [1] believe that SEs should be treated as hybrid organizations that seek to address social problems based on market strategies. The word “hybridity” refers to the use of business models to achieve a profit and nonprofit objectives at the same time. Mair and Marti [2] indicate that SEs aim to create social value. SEs integrate their mission statements into business models and seek to economically sustain themselves to address social issues. SEs, in this light, are not nonprofit organizations but rather organizations in pursuit of both social value and economic value. Therefore, the resolution of social problems is the primary goal of SEs.

SEs are complex and diverse in nature. If we could not adopt a bottom-up approach to understand the conceptual differences and philosophical questions by deciphering the public’s opinions, we would be limiting our research on SEs and entrepreneurship by simply repeating the buzzwords and not exploring the practical applications of the ideas.

Social media has become increasingly important for the communication and information sharing of SEs in recent years. As many SEs are resource constrained, they use free social media for external communications. This is the reason why Twitter is an important tool for trend analysis. Social networks are often the source of big data. Among all social networks, Twitter is among the most important. Twitter has exceeded traditional media in terms of the effectiveness and timeliness of message delivery. Not only private individuals but also governments and corporations like to use Twitter. Therefore, the extraction and analysis of information from social media has become an effective source of knowledge. Tung and Chiu [3] posit that in contrast to large firms, small companies make
an extra effort to use social media. Given the relatively small size of SEs, they cannot afford to miss out on the opportunity of using social media. The use of social media is a superior media strategy for marketing products/services and interacting with customers.

Gallouj and Savon [4] believe that the core of innovative developments or innovations in social services lies in technological change. It is essential that technology be applied to social services to bring about change in the rendering process of social services or in the contents of social services. This paper argues that Twitter is a social and innovative service for SEs. It not only addresses social issues but also creates new social value.

The massive volume of tweets every day makes Twitter an important source of text mining. Tweets are closely related to the surroundings of the people who tweet. Many influential people also make announcements on Twitter. As Twitter generates a large amount of content each second, the analysis of tweets over a time period yields important information or shows the development of events. However, the limitation on the number of words per tweet, the use of abbreviations, erroneous syntax and special signs and characters have been a long-standing problem. These issues have caused problems in text analyzes and searches. Steinskog, Therkelsen and Gambäck [5] highlight the difficulty of topic mining on Twitter given the irregular and unstructured language that is used.

As most data online are unstructured, text mining tools and methods are very important. Text mining refers to the use of content analytical methods to extract information from texts. Different methods are used to analyze textual data and establish connections between words. As text involves unstructured data, it needs to be preprocessed before analysis and the application of statistical techniques or algorithms. The first step of text mining is to analyze various parts of speech. This is followed by the removal of word stems and the recovery of word roots. The screened and selected terms are analyzed for the frequency of occurrence by using statistics or algorithms to identify useful information. The most commonly used method of text analysis is topic modeling, which is an unsupervised classification method used for documents, with the purpose of detecting topics from the corpus. The conventional approach of text analysis typically uses natural classifications for document sets to understand and categorize the texts. In the case of vast, unstructured data or when there is uncertainty about the content, topic modeling can identify natural classifications with the right methodology.

Topic modeling has become a focus of text mining and information retrievals. As technology continues its rapid development, data volumes multiply. Texts are also important data and should not be overlooked, because they contain a large amount of valuable information to be discovered and explored. Social media is at the center of information sharing and socializing in modern life, and includes the most vibrant sharing platforms online. Social media as a force of social change has a far-reaching influence on different aspects of our society. The phenomenon of socialized media is attracting much attention. Facebook, Twitter and other social media see millions of comments being shared by users every day. These comments express the personal opinions of users. Twitter, as an important platform to express thoughts, has become a research domain for text analysis. Twitter’s technology in information retrieval provides new perspectives. In the past, scholars looking into SEs conducted research with qualitative techniques. This paper seeks to apply quantitative methods and perform a statistical analysis on SEs. Meanwhile, extensive literature on SEs focuses on their operations. In contrast, this paper is focused on those who tweet, in an exploration of the factors that influence the behavior of SEs. In fact, there are few studies using latent Dirichlet allocation (LDA) in the research of SEs. This paper aims to establish a new model by devising a new approach to research on SEs. With the collection of big data on social media and the utilization of analytical tools, this paper showcases the advantage of combining media studies and big data analytics. This helps researchers stay on top of the changes and trends in the social media. In sum, the innovation of research methods helps researchers understand the dynamic change of information flows and social media groups.

Twitter provides different features for its social platforms, so that different groups can interact with each other and diffuse information. These different features lead to different data structures.
Twitter limits the number of characters per message because its aim is to quickly distribute information. Due to the rapid generation of big data on Twitter, there is a great deal of noise. It is necessary to filter and preprocess data before content analysis to reduce the possibility of erroneous results. This paper sources a large amount of data via the Twitter Application Programming Interface (API) to demonstrate the potential of big data analytics in contrast with traditional research techniques such as interviews and questionnaires. This is particularly important for research on new areas or domains. This paper suggests the use of text mining as a supplementary method when data collection is difficult. The findings provide evidence of the value of big data and analytics for practical purposes.

This paper summarizes the content, topics and trends of SEs by conducting topic modeling on big data available on social media. Social media is an information platform that evolves over time. The information on social media mirrors changes that occur over time and in the environment. This paper examines the recent development of SEs and identifies the key factors that influence the operation of SEs. The contribution made by this paper is the analysis of content concerning SEs by using data on social media. This paper also explains how topic modeling is used to analyze text contents so that other researchers can have a better framework and template for further studies on SEs. The objectives of this paper are as follows: (1) analyze the contents on the social media outlet Twitter to identify the factors that influence the development of SEs; and (2) develop an index for SEs as a reference for follow-up studies.

The main process used in this paper is illustrated in Figure 1. The process consists of data collection, preprocessing, corpus building, LDA model construction, interpretation and assessment. The first section describes the domain. Section 2 summarizes the literature review. Section 3 outlines the text preprocessing and topic modeling algorithms. Section 4 presents the empirical results. This is followed by evaluations, conclusions and discussions.

2. Literature Review

2.1. Social Enterprise

Defourny and Nyssens [6] indicate that a third sector, encompassing entities such as nonprofit organizations, social cooperatives and mutual societies, provided all kinds of social services in most Western European countries before the Second World War. In the 1950s, the importance of the third sector increased, along with its innovative measures to eliminate poverty and resolve housing problems. In the 1970s, economic recessions and rising unemployment caused the collapse of welfare systems due to fiscal austerity. This was a much larger crisis in the countries used to providing jobless subsidies and pensions. In the 1980s and 1990s, a new type of nonprofit organization slowly emerged as the prototype of social enterprise.

The emergence of SEs and the concept of social entrepreneurship are relevant but different in the European Union and the United States. Defourny and Nyssens [7] believe that the inception of SEs in Europe was primarily driven by social cooperatives in Italy in the 1990s. In 1991, the Italian
government set up a legal framework for social cooperatives. Other European countries followed suit in the late 1990s, in response to the entrepreneurship demonstrated by nonprofit organizations. According to Defourny and Nyssens [6], the EMES European Research Network established in 1996 the definition of SEs: nonprofit but privately owned organizations that offer goods or services to benefit communities, that include stakeholders in the governance mechanism and that emphasize the autonomy of organizational functioning and the economic risks of any undertaking. There are three dimensions in the EMES discourse: economic and entrepreneurial; social; and the specificity of the governance. In other words, the EMES emphasizes the connection between social missions/innovations and products/services, as well as the diversity of stakeholder relations.

Borzaga and Defourny [8] come up with a comprehensive discussion of SEs in Europe and note three theories on SEs. First, the institutional theory considers SEs as an incentive structure. While the wages offered by nonprofit organizations are lower than those offered by for-profit businesses and public sectors, the nonmonetary incentives and high-level freedom unique to SEs can attract and retain employees. Second, the social capital theory formulates a valuable social network via social connections, social systems and social norms and seeks to achieve sustainable economic development with social solidarity. Third, the integrated theory addresses the creation of social value in three aspects, i.e., political, social and economic, within a framework centered on stakeholders.

Meanwhile, SEs in Europe are faced with internal and external limitations. The internal limitations include a lack of awareness of SEs, the trend toward isomorphism, excessively high governance costs and excessively small scales. The external limitations include the underestimation of capabilities held by SEs, the conflict between SEs, the lack of an appropriate legal basis and definite corporate policies for SEs. All these limitations can be improved via management, through the internal/external roles of managers, leadership styles and types of organization.

Defourny and Nyssens [6] believe that the earliest discussion of SEs emerged in the 1960s in the United States. The government invested a large sum of money in nonprofit organizations for the rendering of services in education, healthcare, community development and assistance to those in poverty. Dees and Anderson [9] suggest that there are two schools of thought regarding nonprofit organizations growing in diversity. The social enterprise school emphasizes that the utilization of commercial activities by nonprofit organizations can support and fulfill the mission statements of their organizations. The revenues can create different funding sources for nonprofit organizations. The social innovation school is anchored in the importance, outcomes and social impacts of social entrepreneurs. The functioning of SEs includes innovative services, innovative production methods, innovative organizations or new markets. All schools of thought share the same belief that SEs bring community benefits or create social value. This is the core of mission statements for social entrepreneurs and SEs. Many types of SEs operate in Europe and the Americas. For instance, nonprofit organizations obtain the resources required with commercial means. Alternatively, profit-seeking companies achieve social purposes driven by corporate social responsibility. Galera and Borzaga [10] note three characteristics of SEs: the pursuit of social goals; nonprofit distribution constraints; open and participative governance. Poledrini [11] describes the difference between SEs and other organizations. These include organizational goals, governance, decision-making processes and organizational commitment.

In the U.S., the emphasis is on the integration of social and economic goals; hence, SEs are mainly divided into nonprofit ventures and social investments. In Europe, SEs are about the balance among three entities: society, the economy and policy. Therefore, SEs are mainly in the form of work integration or social innovations. Work integration is the offering of a supportive employment mechanism to resolve long-term jobless problems among disabled and disadvantaged groups. Social innovation focuses on the maximization of social interests via the incorporation of SEs encouraged by laws and the functioning of the market mechanism.

The concept of SE was proposed by the 15 members of the OECD (Organization for Economic Cooperation and Development) and refers to an entity that attempts to achieve specific economic or
social goals through any private activities that have public benefits. SEs do not seek to maximize
profits. SEs are very often associated with social entrepreneurs and social entrepreneurship. Mair
and Marti [2] indicate that these three concepts share the same meaning but have slightly different
perspectives. Social entrepreneurs are the founders of SEs. Social entrepreneurship refers to a process
or behavior.

Young [12] points out two types of SEs: those who contribute to social benefits and nonprofit
organizations seeking to generate a profit by using a business model. Mair and Marti [2] develop
three models for SEs: (1) nonprofit organizations seeking to create social value; (2) profit-seeking
organizations who focus on social responsibility; and (3) SEs seeking to address social problems.
Regardless of the modus operandi chosen, social entrepreneurship has a common core value: nonprofits
that seek to use a business model to create social value. Kerlin [13] posits that SEs use a new method,
combining market-oriented operations and the spirit of nonprofits, to resolve social issues. Dacin,
Dacin and Tracey [14] indicate that the goal of SEs is to pursue public benefits. Folmer et al. [15]
propose that SEs operate with generating financial returns and social value.

In the marketplace, economic goals are the means and presumptions used by SEs to provide a
public benefit. It is essential to generate a profit and ensure the sustainability of the organization.
Perrini et al. [16] believe that SEs need to gain the trust of many stakeholders. This is particularly the
case in the start-up period. A high valuation leads to a greater chance of attracting external capital.
Young SEs may experience difficulties in obtaining external financing due to the limitation of cash flows
and hence need to rely on a diversity of funding resources such as government subsidies, corporate
donations and bank loans. Some scholars argue that social entrepreneurs have distinctive personal
and Canavati [18] mention that in the start-up period of SEs, social entrepreneurs are subject to the
stimulus and influence of two factors: personal achievement drivers and unique business models to
ensure market competitiveness.

Studies on SEs should examine microissues (e.g., social entrepreneurs and organizational
characteristics) as well as macrofactors (such as government agencies, legal systems, economic
conditions, social development and public sophistication). In addition, the effects on SEs differ from
one country to another and across different regions. Huybrechts and Defourny [19] propose a triangular
framework on the basis of fair trade to stimulate thinking in another dimension regarding the economic
structure of SEs. A macroapproach should be taken to examine the influence of politics, education and
advocacy activities. Monroe-White and Zook [20] argue that macrosystems affect innovation factors
such as the products, marketing and business models of SEs. Ghods [21] identifies the four elements of
entrepreneurial marketing for SEs: market competition, capital access, volunteer recruitment and the
provision of products/services for targeted audiences. SEs are a social and innovative industry against
the backdrop of social changes.

While we seek to honor our mission statements and meet the demands of the disadvantaged by
deploying innovative methods, creating social value or distributing public goods, we should also take
heed of the variances across regions, cultures and property attributes regarding the environment or
systems in support of innovations. This will help align innovations with practical requirements.

Hazel and Onaga [22] posit that the characteristics of all participants, the features of the
organization, the possible interactions and effects of the community or the environment and the
government’s support for innovative ideas are all potential key success factors of social innovations.
Mulgan [23] indicates that nonprofit organizations are often subsidized by the government. While they
help governments offer many kinds of social welfare services, their innovative ideas are not acceptable
to governments. Therefore, the timing and key success factors of social innovations depend on whether
such social innovations trigger social changes and drivers for continued social development. Djella
and Gallouj [24] contend that social innovations are not focused only on innovation and trends in
social and economic organization or the third sectors. They are also focused on innovation and
changes in social services. The utilization of social innovations should take into consideration the
social, environmental, economic, political and local cultural contexts. Many strategies of social innovations have been adopted around the world. Nicholls and Murdock [25] note the observable aspects of social innovations. These innovations are observed at four levels, including processes and outcome. They also categorize social innovations into eight types, such as community environments or living standard betterments. Muhammad Yunus founded Grameen Bank to offer microloans to help poor people improve their livelihood. In sum, government intervention can have major implications for social innovations. Partnering with the government can resolve more social problems. Lyon [26] posits that social innovations reflect the development of a concept in the resolution of social problems. The success of social innovations depends on the collaborative relations of SEs. Additionally, the establishment of social networks and social connections is the key for working with SEs. Both funding from the government and the environment of social institutions should be assessed and managed. Tortia, Degavre and Poledrini [27] believe that any knowledge or ideas from members of SEs driven by personal motivations or management decisions within SEs in response to social changes are a product of social innovations. Social motivations, collective actions, stakeholder governance and social resource integration are the main drivers of innovations. Meanwhile, the success of social innovations depends on the existence of a mature and stable market to assist the functioning of organizations. Social innovations will be hard pressed if they cannot lead to long-term and stable changes to mainstream systems or concepts.

Social innovations are meant to satisfy social needs with new ideas and approaches, so that innovative activities and services can be widely promoted and developed. The value of social innovations consists in fresh thinking and new solutions, resulting in new and consensual social values. Therefore, a social innovation can address social needs more quickly and effectively and create new social relations and social value networks. Social innovations may occur whenever society is faced with new drivers and challenges as a result of demographic shifts, social evolutions, technological transformations, new demand paradigms or cost increases. When confronting these trends and challenges, social innovators need to ponder new service contents, new management styles and new resource allocations to maintain the sustainable development of SEs.

2.2. Topic Modeling

When mining unstructured texts, machines should first be used to comprehend the language, which is called NLP (natural language processing). All the topic analysis methods are based on NLP. The first step is tokenization, which allows machines to understand each word, so that they can understand the sentences and analyze their meaning, grammar and syntax. The unigram model and a mixture of unigram models were developed by Nigam, McCallum, Thrun and Mitchell [28], and LSA (latent semantic analysis), which was proposed by Deerwester, Dumais, Furnas, Landauer and Harshman [29], can be used for this purpose. However, topic modeling is a mixed membership model, different from the unigram model or a mixture of unigram models. The unigram model assumes that each word has the same term distribution. In contrast, the mixture of unigram models assumes that each document contains multiple topics. All the words form different terms and are distributed into different topics in the same document. However, mixed membership models assume that each document contains multiple topics and that the same word may appear in different documents.

Term frequency–inverse document frequency (TF-IDF) is a technique frequently used in topic modeling for searching and text mining, to assess the level of importance of each word in a document or corpus and hence select the words that distinguish the document. There is a positive correlation between the frequency and importance of a word. However, if the same word appears frequently in different documents in the corpus, the importance of this word is reduced, given its lack of discrimination for different documents.

The most frequently used algorithm for topic analysis is LDA, which is developed by Blei, Ng and Jordan [30]. This is the Bayesian mixture model based on the extension of the probabilistic latent
semantic indexing (pLSI). This method considers documents as a probabilistic model suitable for dealing with segmented data with no connections between topics.

The LDA model contains three layers: terms, topics and documents. Each topic is a representation of the probabilistic distribution of the terms relevant to the topic concerned. LDA can identify information on a topic in a large set of files or corpora. In other words, LDA deals with unstructured texts by grouping the vocabulary terms distributed in a corpus into different topics and synthesizing these topics into documents. LDA can be used to pinpoint the blending of relevant words for each topic and the constitution of topics in each document. Hong and Davison [31] suggest that the generalization of each document consists of three processes: (1) the random selection of a topic from many topics in the document; (2) the selection of words relevant to the topic chosen; and (3) rinsing and repeating the above processes for all the terms in the document until each topic is clarified.

SEs are the new reformers in society and the corporate world. They strive to be proactively involved in public affairs and service delivery to resolve the crisis of welfare states. This proves the importance of studies on SEs. This paper analyzes the contents on the social media outlet Twitter to identify the influencing factors of SE development. The development of topic modeling started from the vector space model, in the early days, and led to the LDA topic models, with both structure and functionality improving over time. Based on the innovative application of model structures, this paper analyzes the meanings behind tweets about SEs and develops models with optimization algorithms. Given the irregularity of topics that appear on Twitter and the brevity of tweets, topic modeling is a challenge. However, Weng, Lim, Jiang and He [32] prove that LDA is effective for tweets. Many scholars apply LDA to different domains, such as transportation and politics. Sun and Yin [33] employ the LDA method for topic modeling traffic information and contribute by proposing new definitions to be used in academic research. Yoon, Kim, Kim and Song [34] deploy the LDA technique to identify topics related to Korean politics on Twitter. Meanwhile, the use of LDA is also popular in the health field [35].

3. Materials and Methods

The corpus in this paper consists of the tweets extracted from the Twitter API. The data are unstructured texts. The LDA method is used to explore potential subdomains of SEs. Below is an explanation of the data sources and empirical methodology used in this paper.

3.1. Data Collection and Preprocessing

First, we acquired older tweets after an application was submitted to access Twitter’s Developer Premium API. This was followed by the establishment of the interface with the API by using the “create_token” function, a package in R language. Then, the search_fullarchive function was used by searching for the #SocEnt hashtag in tweets in the English language from June 2018 to September 2019. The sampling pool consisted of 232,786 tweets.

The tweets include upper/lower case letters, changes in the parts of speech and many redundant phrases. Therefore, this paper uses the embedded function “tolower” in R language to convert all the texts into lower case. This was followed by the recovery of parts of speech by using the wordStem function in SnowballC. The tidytext and tidyverse packages and regular expression were then deployed to delete stop words and meaningless symbols such as URLs, usernames, punctuation, numbers and outliers. These elements serve little value in practice. Then, stemming was employed to trim off variations and repeated jargon and restore the word roots. Finally, 2,454,080 distinctive words were tokenized. This study does not determine the importance of individual words using TF-IDF because tweets are short texts. Empirical studies on natural language processing suggest that TF-IDF is not suitable for processing words for the purpose of distinguishing documents.

3.2. Topic Modeling Using LDA

LDA works by exploring the potential structure of texts and the finalized collection of topics and involves a process of synthesizing the topic structures and documents (θ, z) based on the words (ω)
observed. Blei, Ng, and Jordan ([30], p997) illustrates the LDA model. The $M$ denotes the total number of documents, and $N$ denotes the total number of words in a document. The first level represents the parameters of the corpus. In this paper, the $\alpha$ and $\beta$ indicate the parameters sampled, $\theta$ represents the variable sampled once in each document, and $z$ and $\omega$ are the variables representing each word sampled one time in each document. Before conducting the LDA analysis, it is necessary to assign a fixed $K$ value as the number of topic models. The LDA model assumes that the process of document generation in the corpus can be expressed as $w = (w_1, \ldots, w_N)$. A vocabulary of $N$ words can be decomposed into $V$ terms; for example, $w_i \in \{1, \ldots, V\}, i = 1, 2 \ldots, N$. Therefore, the LDA generation model uses three steps [36].

4. Results

4.1. Per-Topic Word Distribution

The first step of topic modeling is preprocessing to delete unnecessary characters or signs such as punctuation, emoticons, Twitter’s specific signs, numbers, URLs and stop words. Overall, a total of 2,454,080 words were obtained from the corpus. This step is followed by frequency screening to identify the most frequently used words in the source. This paper samples the top 50 words that appear more than 300 times. Figure 2 provides the distribution of each word and the connections between words. The bottom of Figure 2 shows a group of words with tighter connections with each other. Section 4.2 lists the different topics identified based on these connections.

![Figure 2. TM Twitter Data Using the Hashtag: SocEnt.](image)

Figure 3 shows the ranking of the words according to frequency and importance. The top 10 words with the highest frequency are social, impact, busi (business), support, communiti (community), join, people, enterprise, learn and entrepreneur. These words are then visualized into word clouds (Figure 4). According to the results shown in Figures 3 and 4, this paper derives two findings that are in line with previous studies. (1) SEs strive to convey social value. This is evidenced by the use of keywords such as “social, impact, business, community and enterprise”. (2) Such social value needs to be conveyed to
more people. Whether latecomers and followers join the bandwagon is a key determinant of the success of SEs, evidenced by the use of keywords such as “support, join, people, learn and entrepreneur”.

![Figure 3. Top 50 Most Salient Terms.](image1)

![Figure 4. Word Clouds of Social Enterprises (SEs) on Twitter.](image2)

4.2. Topic Modeling and LDA Visualization

This paper uses the *topicmodels* package in the LDA function written in R language. This requires the determination of the K value, which is the number of topic models generated from the corpus. To decide the K value in an objective way, this paper adopts the selection method proposed by Deveaud [37], which obtains the K value using the `FindTopicsNumber` function of the *ldatuning* package.
package. Figure 5 shows that the optimal K value is 3. Therefore, this paper chooses three topics for the LDA model (k = 3). The terms generated for each topic are presented in Table 1.

![Figure 5. Number of Topics (K).](image)

**Table 1. Terms Listing of Topics.**

<table>
<thead>
<tr>
<th>Topics</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>social, people, free, world, business, event, enterprise, support, share, programme, future, starting, entrepreneur, innovation, community, pour, impact, accelerator, book, food, awards, create, life, products, fantastic, start, meet, website, win, exciting, change, employment, amazing, challenges, idea, city, love, award, workshop, funding, youth, sustainable, based, building, grow, delighted, story, ambitious, chocolate</td>
</tr>
<tr>
<td>Topic 2</td>
<td>impact, join, people, learn, world, change, support, team, social, amazing, network, local, opportunities, scale, opportunity, women, entrepreneurs, sector, start, space, learning, food, enterprise, global, news, helping, positive, community, growing, event, free, economy, leaders, funding, meet, ready, access, programme, meeting, movement, shop, sustainable, tomorrow, happy, power, supporting, excited, innovation, proud, love</td>
</tr>
<tr>
<td>Topic 3</td>
<td>social, business, apply, impact, support, community, entrepreneurs, join, enterprise, investment, entrepreneurship, businesses, communities, programme, people, program, women, fund, ideas, development, excited, team, create, applications, creating, founder, session, skills, proud, register, ceo, cohort, hear, investing, jobs, mit, building, change, launch, training, award, learn, scale, global, sharing, money, financial, details, accelerator, story</td>
</tr>
</tbody>
</table>

**Source:** Authors' own statistical analysis.

To better differentiate the relations that exist between the topics, this paper produces a visualization of the LDA analytical results by performing the **LDAvis** function written in R language. **LDAvis** is a web-based platform used for interactivity and visualization by using **D3.js**. Figure 6 presents the spatial distance between the three topics and the interactions among the top 30 terms contained in each topic. **LDAvis** is a tool used to explore the relations between the topics and terms. This figure has two panels. The pie charts on the left present the distribution and size of each topic in a two-dimensional space. If there is no interaction between two topics, it means they are mutually independent. If there is an interaction, then a connection exists. The size of each pie chart represents the number of articles on the topic. The right panel provides bar charts showing the 30 words most frequently used (in red) to discuss a specific topic and the total frequency (in light blue) of individual words in the corpus.

The optimal parameter value in this paper is K = 3 (the number of topics). The left side of the graph in Figure 6 shows three pie charts. Each pie chart indicates a topic. Inside each pie are the 30 most frequently used words for the topic. These words are arranged from the highest frequency to the lowest frequency on the right side of the graph. The graphs represent Topic 1, Topic 2 and Topic 3, respectively. The pie charts on the left are independent, without any overlap. This means that all three subjects are independent of each other. The right side of the graph shows the frequencies...
of individual words in the Twitter corpus and the percentage of such frequencies in a single topic (the percentages of red and light blue). These percentages are very meaningful, as they indicate the level of importance of a word in the documents, expressed as TF-IDF in the literature. Words with discriminating power are selected for different documents. If a word appears frequently (measured in percentages) in a topic, it has discriminating power for that topic, and it is very meaningful to that topic. However, if a word appears frequently in a topic but also in other topics, it means that the word is very important but does not have discriminating power over a single topic. The word “social” appears with the highest frequency for the topic shown in Topic 1 of Figure 6. However, the word “social” also appears frequently in the other two topics. Therefore, it is not representative of Topic 1.

Figure 6. Cont.
The darker the color, the greater the importance of these words in the topics. The words in light yellow report a high frequency (among the top 30) in each subject, but they are less important compared to other words in the same topic. This is the reason for their lighter shade.

To better illustrate the importance of the terms used for discussing each topic, this paper presents a visualization of these terms with word clouds plotted using the `lda.plot.wordcloud` function in `corpustools` in R language. Some bar charts presenting the frequency of the words are produced with the LDAvis package. However, word clouds are a more intuitive presentation of frequencies because they are indicated using size and color depth to help capture the core vocabulary of each topic. Figure 7 shows the word clouds of the three focal topics.

The word clouds shown in Figure 7 present the frequency and importance of words in different topics with font sizes and color shades. The larger the font size, the higher the frequency of appearance. The darker the color, the greater the importance of these words in the topics. The words in light yellow report a high frequency (among the top 30) in each subject, but they are less important compared to other words in the same topic. This is the reason for their lighter shade.
As shown in Figures 6 and 7, the most frequently used words in Topic 1 are “world”, “support”, “future”, “entrepreneur” and “innovation”. These words in aggregate suggest that the success of SEs...
requires an effective innovation strategy. Everything else needs to work in sync as well to enhance the likelihood of success. This includes policy support, human resource management and corporate sponsorships. This paper defines “strategy” as Topic 1. The most frequently used words in Topic 2 are “impact”, “join”, “people”, “learn” and “change”. The word “impact” has the highest frequency of appearance. Impact is a company’s contribution to society via its product or business operation. Impact is manifested with positive effects on society and the value desired to be created. It is possible to connect with the vision and mission stage of SEs via the Theory of Change (TOC). Simply put, TOC indicates different changes in specific conditions or situations. This is the reason why this paper refers to the word “impact” to represent Topic 2. The most frequently used words in Topic 3 are “business”, “apply”, “support”, “community”, “enterprise” and “investment”. The word “business” is the word with the highest frequency of appearance, only second to the word “social”. This demonstrates the unique value of SEs via double bottom lines, i.e., SEs, social value and economic value. Both social goals and business targets are essential. SEs need capital from investors, as well as government resources and funding support. An investment case is the basis of community development. Therefore, this paper uses the word “business” to represent Topic 3.

4.3. Topic Analysis

This paper presents a visualization of the key words by using the LDAvis function and then analyzes the word clouds. LDA is a commonly used method for topic modeling. LDA is used to identify a mixture of relevant words used for each topic and the topics that appear in each document. A word may have different degrees of importance for different topics. For instance, the word “people” may appear in different word clouds but at different frequencies and with a different degree of importance. Table 2 summarizes the three focal topics based on the most frequently used words in each word cloud.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Topic Label</th>
<th>Top 10 Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>Social strategy</td>
<td>social, people, free, world, business, event, enterprise, support, share, programme impact, join, people, learn, world, change, support, team, social, amazing</td>
</tr>
<tr>
<td>Topic 2</td>
<td>Social impact</td>
<td>social, business, apply, impact, support, community, entrepreneurs, join, enterprise, investment</td>
</tr>
<tr>
<td>Topic 3</td>
<td>Social business</td>
<td>Source: Authors’ own statistical analysis.</td>
</tr>
</tbody>
</table>

Social strategy (Topic 1): SEs exist to solve social problems. Social strategies are used when a social entrepreneur identifies a social issue and seeks to resolve it by establishing a social enterprise. Social entrepreneurs act on their ideas and motives, as well as their insight and new perspectives of social value. Peredo and McLean [38] indicate social entrepreneurship as the actions taken by some people or groups that involve (1) the creation of social value as the goal; (2) the ability to identify good opportunities; (3) the use of innovation and ideas; (4) the willingness to face certain risks; and (5) the confidence to deal with situations with existing and limited resources. In summary, the operating strategy and social creation are two sides of the same coin for SEs. Value is created in a business model and structure that involve all the participants in the value network and are built on products, service and information flows [39]. The dimensions of value creation include key resources, key workflows, main value activities and network partners [40–42].

Social impact (Topic 2): A social enterprise begins to exercise its influence the moment it starts to function. The ultimate outcome of such influence is to change people. During the process, SEs consider all stakeholders and successfully convey their value to all parties [43]. The maximization of shareholder value is by no means the main mission of SEs. However, stakeholders are of great importance. Generally, SEs are created to have a specific impact.
Sustainability (Topic 3): Once SEs have established their strategies and social purposes, the next step is to formulate the methods to achieve the desired benefits and convey the social value, which involves the design of a business model. The customers, suppliers and strategic partners that are involved in the business structure exchange monetary value, as well as knowledge and intangible benefits (e.g., customer loyalty, awareness and image). This goes beyond the focus on products, services and revenues [44]. The value creation process is essential for SEs. The maximization of shareholder value is not the primary objective of SEs. Mair and Schoen [45] indicate that SEs are different from mainstream corporations in terms of the value delivery process. This is because SEs integrate their customers and target audiences into the value creation process. The central concept of SEs is different from that of regular companies. If the characteristics of an element are changed in the process, the interaction and procedural relations between elements also change. This is the reason why the business models used by SEs are changeable and ambiguous. More specific definitions are required to understand these models [46].

Each of the three topics (strategy, impact and business) derived by this paper contains a rich collection of words. While the word “people” appears in all three topics, it is not adequate to distinguish the topics using topic modeling. However, this word is frequently used for all topics, which indicates its importance. Hence, this study incorporates the “people” factor because SEs exist to solve people’s problems in society and need to convey the social value created to people. In other words, people are an important element of the business models used by SEs. Identifying people willing to accept change or change on their own to adhere to the value propositions of SEs is the key to the sustainability of SEs.

This study not only identifies the abovementioned three topics and four dimensions but also separates the analytical findings regarding the words that appear in the corpus into six indexes: S, O, C, E, N and T. This process also involves the initial search for the hashtag #SocEnt, which is ideal when considering the framework of SEs and for providing a template for the development of SEs going forward. All the possible patterns, business models, social values or impacts of SEs can be interpreted and analyzed by using these six indicators. These six indexes, SOCENT, can represent the entire spectrum of SEs. The core meaning of each index is described in the following:

Social: The term “SEs” speaks for itself. SEs are anchored in society, as they seek to resolve social problems and create social value. Many words used to discuss the topics are about this concept.

Opportunity: Because SEs endeavor to solve social problems, they need to identify good opportunities that can act as a starting point for addressing external environments, government support, legal frameworks, business development, economic changes, cultural backgrounds, acceptance and participation by the public. All of these factors can provide opportunities for the first step of SEs.

Change: SEs seek to address social issues and communicate their social value. In other words, they want to change society. It is important for SEs to specify the desired change. In brief, SEs must articulate the value proposition, objectives and effects.

Enterprise: Social value provides the foundation of SEs. A good business model will consistently convey this value. A good business model is required so it can maintain the functioning of SEs and continue to communicate their social value.

Network: Society consists of groups of people. SEs cannot stand on their own, apart from organizations or entities, or complete social change single-handedly, without support or a supply chain. The use of social resources for greater benefits is a key metric used by SEs. In summary, the networks are “social networks”. Folmer et al. [15] emphasize that SEs obtain more intangible resources from networks than commercial companies. Kokko [47] posits that the network relations of different stakeholders contribute to the various developments and outcomes of SEs. Therefore, different types of social value are provided by SEs. This means that the economic model used to value firm profits cannot be used to assess the success of SEs.

Team: Teams are made up of people. Team members who support SEs are key to their success. Social entrepreneurs organize teams, gradually develop organizations and eventually influence all the people involved. This is an extension of the “team” concept.
Based on the literature review and topic modeling results, this paper inducts and deducts three subjects and four dimensions. The most frequent 50 words are screened from the corpus. Then, in-depth interviews with social entrepreneurs are performed to compare and validate the key factors. This is followed by the generalization of six indicators as the basis for the construction of the theoretical framework. These six indicators are used to interpret the three subjects, to report a complete dynamic process of SEs.

5. Discussion

The emergence of social media allows SEs to build their networks, promote social developments and hence have social influence. In brief, social media is a platform used to convey the core value of SEs. Schwartz [48] believes that the creation, communication and delivery of value cause consumers to focus on social marketing. Lawrence [49] indicates that social media allows users to make connections, have conversations and cooperate and can also generate the interest of new users.

Social media is a highly influential channel of digital communication. SEs can establish connections with target audiences via social media, which will cause individuals to care about social issues and create value. SEs share and broadcast their mission statements via social media to promote their social impact and communicate their value. This allows trusting relationships to be formed and promotes value creation. Given the limited resources available, SEs should place extra emphasis on the potential benefits of social media.

This paper conducts a literature review on SEs and social innovations in Europe and the U.S. to establish research topics by taking into account both objective and subjective factors and generalizing the common traits of SEs. The establishment of big data analytics helps to clarify the concept of SEs.

The findings of this study are consistent with those of previous studies. For instance, tweets can be used to reveal the multiple dimensions of SEs [1,2]. Texts about the economic value and social value of SEs appear in tweets [15,39] and can be separated into main topics. This suggests that the business models and social impacts of SEs evolve along with social developments, and as a result, SEs come in different forms and have different patterns. The six indicators proposed by this paper are in line with the macro- and microperspectives taken in previous studies [19,20]. The terms social, opportunity and change can be meaningful for the use of macrometrics. The terms enterprises, teams and networks can be used for micrometrics that may shed light on possible directions for further studies.

Based on the optimal K criteria and LDA algorithms, this study identifies three topics that arise for the Twitter hashtag #SocEnt. These three topics are social strategy, social impacts and social businesses. The word clouds for each of these three topics are extended into four dimensions (strategy, impact, business and people). The words in the corpus are synthesized into six indexes for SEs. The recent expansion of SEs can be summarized with the theoretical framework depicted in Figure 8. This framework can serve as a template for future and further studies on SEs.

Figure 8 depicts the theoretical framework of SEs developed by this paper. It contains three levels, i.e., elements, key operating factors and benefits, described as follows:

Elements (level 1): The elements of SEs include social strategies, social impacts and social businesses. This is a very important framework for the thinking of social entrepreneurs. SEs exist to resolve problems and create change. In the midst of any social issue, the first step is to develop a strategy for problem solving. This strategic thinking then leads to the business model of SEs and the eventual impact to drive social change.

Key operating factors (level 2): There are six indexes covering the development process of the business model for a social enterprise. These indexes are social, opportunity, change, enterprises, networks and teams. The relationship between these six indexes and the elements of SEs is as follows.

1. **Strategy formation and business models**: Strategic thinking for SEs must take into consideration the opportunities in the external environment for decisions over business models. Only with good commercial opportunities and corporate operations can a social enterprise achieve sustainable development (index: opportunity, enterprise).
(2) **Business models and social impacts:** The purpose of SEs is to create social impacts and social changes. This makes the core value and philosophy of the entrepreneurs and teams critical to the operational process, as it is necessary to change and adjust business models according to the status of the society (index: team, change).

(3) **Social impacts and strategy formation:** To create social influence, SEs must have a good social network to expand social benefits. The improvement in social problems and the change in values are often the basis for determining whether a social enterprise is successful or not (index: social, network).

Benefits and assessments (level 3): In the process of resolving problems and creating impacts, SEs must retain their own value creation and strengthen the delivery of value networks. The speed of conveying and broadcasting its value is often the key performance indicator of a social enterprise. The formation of a strategy is necessary to create the value that it wishes to deliver, and the success of a business model is required for the social diffusion of such value.

Twitter provides a very large amount of content that can be used by researchers. The use of text mining has become popular in various fields. However, researchers will find the analysis of the unstructured and short texts posted on Twitter a daunting challenge. In this paper, the posts on social media are searched and the topics related to SEs are classified. This paper is important because it revisits and analyzes the core value of SEs using big data.

![Figure 8. Framework of SEs.](image)

6. Conclusions

The big data techniques used by this paper can be applied in research on other topics. The research methodology used here can serve as a template for future studies. This paper examines topics related to SEs using the optimal K value (i.e., the number of topics), which is 3. The research findings highlight the limitations of the development of SEs. The literature suggests that SEs use sustainable business models and exercise their social influence. Even more important for the development of SEs is government support, legal restrictions, changes in the social environment and the public’s acceptance...
and involvement. There is a long list of complicating and intervening factors, including social aspects, policy measures, economic status, environmental conditions and appropriate timing. As a result, the long-term development of SEs is constrained. Going forward, SEs should seek to overcome these hurdles to social innovation.

The text classification techniques used in this paper can be used to address the difficulty of analyzing short texts (such as tweets or other social media content). This paper employs these techniques to analyze topics related to SEs. It is hoped that future studies can extend research on the scope of SEs by considering trending texts on other social media and news platforms. This will help to clarify and crystallize the makeup of SEs.

However, obtaining the API authorization from Twitter represents a challenge that will be faced during the data collection process. Twitter limits the access to data on its API, especially historical data. If data are to be collected via the API, the best approach is to collate information over an extended time period. Otherwise, it will be difficult to acquire sufficient historical data, and thus the research analysis and results will be limited. A generous budget should be acquired prior to beginning the study because it will be difficult to access data via the Twitter API if resources are limited.

**Author Contributions:** Conceptualization, S.-W.K.; methodology, S.-W.K.; software, S.-W.K.; validation, S.-W.K.; formal analysis, S.-W.K.; investigation, S.-W.K.; resources, S.-W.K.; data curation, S.-W.K.; writing—original draft preparation, S.-W.K.; writing, review and editing, S.-W.K. and P.L.; visualization, S.-W.K.; supervision, P.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors are most grateful for the helpful comments and suggestions of the editor and three anonymous reviewers.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


49. Lawrence, A. Is SCRM really a good social media strategy. *J. Database Mark.* 2011, 18, 149–153.

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