ABSTRACT
The purpose of this study is to provide a bibliometric analysis of customer engagement (CE) research in the period 2006-2021 by using the PRISMA protocol for systematic reviews and by relying on a set of CE-related keywords. Bibliometric analysis refers to the quantitative study of bibliographic material that provides a general picture of a research field. By using a bibliometric analysis, the most relevant research in a particular field can be provided and the newest research trends can be identified. This study will provide a detailed overview of the evolution of relevant literature and the status of CE research over the past 15 years by using VOSviewer software for creating, visualizing, and exploring bibliometric maps of science. The concept of CE emerged in the marketing literature around 2005 followed by an increased number of research conducted in various contexts and fields, from customer and firm perspectives, etc., linking customer engagement to different marketing concepts such as customer satisfaction. Additionally, some of the researchers conceptualized customer engagement as a behavioral concept whereas others conceptualized it as a psychological concept. Based on the need for further clarification of this concept, a systematic review through bibliometric analysis was conducted and the results of descriptive analysis (distribution of articles by year, top five journals based on the number of published articles, top ten most cited articles, and country co-authorship network visualization) are presented. Additionally, the results from keywords co-occurrence analysis based on text mining in the abstracts are shown. Moreover, a machine learning algorithm for logistic regression in Power BI Desktop was performed to identify independent variables associated with greater citations of CE research. The results of the performed bibliometric analysis can be used by marketing scholars as a basis for future CE research.

Keywords: bibliometric analysis, customer engagement, keyword co-occurrence analysis, mapping

JEL classification: M30
1. INTRODUCTION

In contemporary marketing thought, the role of customers is becoming continuously more important due to their participation and collaboration in companies’ marketing activities, such as brand communication (Hamilton et al., 2016), product referrals (Van Doorn et al., 2010), product development (Hoyer et al., 2010), and mutual support, in general. In fact, various customer behaviors beyond transactions (Van Doorn et al., 2010; Brodie et al., 2011) refer to customer engagement (CE) which is emerging as an important determinant of firm performance as both tangible and intangible outcomes (Pansari and Kumar, 2017).

The concept of customer engagement emerged in the marketing literature around 2005 (Sawhney et al., 2005) and since then its academic prominence is continuously rising along with its practical proliferation. Since 2010 the Marketing Science Institute has indicated CE as a key research priority (MSI, 2010, 2020) and during the last decade, the academic interest in this concept has been escalating (Hollebeek et al., 2022). In spite of the extensive scholarly debate, there is a lack of agreement on how to conceptualize and operationalize this concept (Lim et al., 2022; Syrdal and Briggs, 2018). Defined broadly, CE is understood as a psychological state (Brodie et al., 2011) or sequential psychological process (Bowden, 2009) related to customer-brand interactions (Hollebeek, 2011) which refers to “specific levels of cognitive, emotional, and behavioral activity” (Hollebeek, 2011, p. 790). On the other hand, Van Doorn et al. (2010) applied a behavioral perspective in comprehending and defining CE, and a vast majority of studies have analyzed various forms of behavioral manifestations (Jaakkola and Alexander, 2014; Pichler et al., 2019).

Having in mind the theoretical fragmentation of this concept (Hollebeek et al., 2019; Hollebeek et al., 2022) and the reached maturity for a review to be conducted (Brodie et al., 2011; Hollebeek et al., 2019; Hollebeek et al., 2022; Lim et al., 2022), this paper strives to provide a systematic mapping of CE by conducting a bibliometric analysis and keywords co-occurrence analysis followed by an analysis of independent variables associated with greater citations of CE research. More precisely, relying on a set of CE-related keywords and based on the PRISMA protocol, the CE research in the period 2006-2021 is reviewed by providing a descriptive overview in terms of distribution of articles by year, top five journals based on the number of published articles, top ten most cited articles, and country network visualization, as well as keywords co-occurrence analysis to reveal the most prominent research topics. In this line and based on the developed questions in the previous review studies on CE (Rosado-Pinto and Loureiro, 2020, So et al., 2021; Srivastava and Sivaramakrishnan, 2021; Hollebeek et al., 2022; Lim et al., 2022), the following research questions are proposed:

RQ1: What are the bibliometric trends (performance) of CE research?
RQ2: What are the top researched words and topics in CE research?
RQ3: Which are the independent variables associated with greater citations of CE research?

Although recently some review studies on customer engagement are evident (Rosado-Pinto and Loureiro, 2020, So et al., 2021; Srivastava and Sivaramakrishnan, 2021; Hollebeek et al., 2022; Lim et al., 2022) and they also quantitatively explored the intellectual structure of CE and its trends, the present study is the first that goes further in exploring the independent variables associated with greater citations of CE research by applying machine learning algorithm for logistic regression. In addition, the present paper covers a broader research scope by comprising several related keywords to customer engagement, i.e., consumer engagement, consumer brand engagement, customer brand engagement, customer engagement behavior/behaviour, consumer engagement behavior/behaviour, and brand community engagement. Comprising more than one keyword as well as covering all the
journals (not only ranked ones) resulted in a total number of 2201 reviewed studies which is the most exhaustive number of studies compared to the previous reviews on CE. The remainder of the paper proceeds as follows. First, the paper provides a theoretical overview of the CE concept. Next, the methodology and data are presented. Finally, the results and analysis are elaborated, and future research is discussed.

2. THEORETICAL BACKGROUND

The concept of engagement has been analyzed from different aspects in different fields such as psychology, sociology, organizational behavior/management, etc. Within the engagement concept, studies have investigated social engagement (Krueger et al., 2009; Bath and Deeg, 2005), student engagement (Trowler, 2010; Carini et al., 2006), civic engagement (Galston, 2007; Youniss et al., 2002), employee engagement (Saks and Gruman, 2014; Anitha, 2014), stakeholder engagement (Greenwood, 2007; Maak, 2007), etc. thus providing valuable insights into the concept beyond the marketing discipline. Marketing scholars have begun to explore the engagement concept by focusing on CE at the beginning of the new millennium, and since then CE has been broadly investigated in the marketing literature which is reflected in the increased number of published CE papers in reputable scientific journals (Bowden, 2009; Van Doorn et al., 2010; Brodie et al., 2011, Hollebeek et al, 2022). This increased academic interest is related to the increased importance of the CE topic among practitioners having in mind the changes in the business environment with regard to digitalization and increased use of the internet and social media by both customers and companies.

However, when analyzing the CE research it can be noticed that the CE articles rely on different theoretical perspectives (Hollebeek et al., 2019; Hollebeek et al., 2022; Ng et al., 2020), they focus on different conceptualizations of CE (Hollebeek, 2011; Van Doorn et al., 2010; Bowden, 2009), analyzed CE from the firm (Kumar et al., 2010; Meire et al., 2019) and customer perspectives (Brodie et al., 2011; Kumar et al., 2019), analyzed CE in different contexts (Srivastava and Sivaramakrishnan, 2022; So et al., 2020; Hao, 2020; Opute et al., 2020; Chen et al., 2021) and provide a variety of definitions (Harmeling et al. 2017). Besides different conceptualizations of CE, the interpretations of the concept of CE differ among the practitioners as well (Sashi, 2012).

Regarding the conceptualizations of CE, CE has been analyzed from the behavioral aspect (Van Doorn et al., 2010), and as a “psychological state” (Brodie et al., 2011; Hollebeek, 2011) whereas Bowden (2009) describes CE as a “psychological process”. Consequently, different definitions of CE have been offered. Thus, according to Van Doorn et al. (2010, p. 254) customer engagement is defined as “a customer’s behavioral manifestations that have a brand or firm focus, beyond purchase, resulting from motivational drivers”. Brodie et al. (2011, p. 260) considered CE as a psychological state that “occurs by virtue of interactive, co-creative customer experiences with a focal agent/object (e.g., a brand) in focal service relationships.”. On the other hand, Bowden (2009) defined customer engagement as a sequential psychological process that individuals move through to become loyal to the brand, whereas Hollebeek (2011, p. 790) defined CE as “the level of an individual customer’s motivational, brand-related, and context-dependent state of mind characterized by specific levels of cognitive, emotional, and behavioral activity in direct brand interactions.”. Moreover, there are differences in the research articles regarding analyzing the CE construct as unidimensional (Verhoef et al., 2010; Shen et al., 2019) or multi-dimensional. However, most of the CE studies analyzed the CE construct as multidimensional and most of them as a three-dimensional (cognitive, emotional, and behavioral) concept (Brodie et al., 2011; Hollebeek, 2011; Vivek et al., 2012; Hollebeek et al., 2014; Dessart et al., 2015).
Furthermore, most of the published CE articles mainly study CE in relation to brands, with customers being the central engagement subject (Hollebeek, 2011; Hollebeek et al., 2014; Brodie et al., 2011) although there are studies that focus on consumers as well (Sharma et al., 2021; Hollebeek et al., 2014; Rather, 2019). In marketing-related research studies, it can be noticed several related terms to customer engagement (Verhoef et al., 2010; Brodie et al., 2013; Bowden, 2009), such as consumer engagement (Brodie et al., 2011; Vivek et al., 2012; Dessart et al., 2015), customer/consumer engagement behavior/s (Van Doorn et al., 2010; Jaakkola and Alexander, 2014; Oh et al., 2017), customer/consumer brand engagement (Van Doorn et al., 2010; Hollebeek, 2011; Hollebeek et al., 2014); brand community engagement (Baldus et al., 2015; Kumar and Kumar, 2020), etc. Given that customer engagement and consumer engagement are mutually related concepts (Hollebeek et al., 2022) and that CE mostly has a brand focus (Van Doorn et al., 2010; Sharma et al., 2021; Kumar and Kumar, 2020) our review study relies on several related terms (customer engagement, consumer engagement, customer brand engagement, customer brand engagement, customer engagement behavior/behaviour, consumer engagement behavior/behaviour, and brand community engagement), unlike the other studies that analyzed a single or few keywords.

The increased number of CE research studies led to increased research interest in review studies on CE. The first review study on CE in marketing is the study of Islam and Rahman (2016) and since then, the research attention on reviews of customer engagement empirical research has been continually increasing (Rosado-Pinto and Loureiro, 2020; Rasool et al., 2020; Hao, 2020; Hollebeek et al., 2022; So et al., 2021). However, although the recent but limited review studies used quantitative analysis (Rosado-Pinto and Loureiro, 2020, Hollebeek et al., 2022; So et al., 2021; Srivastava and Sivaramakrishnan, 2021; Lim et al., 2022) none of them used regression analysis.

3. METHODOLOGY AND DATA
For conducting a transparent systematic literature review we followed all four stages of the PRISMA protocol: identification, screening, eligibility, and inclusion (Moher et al., 2009). First, by searching the SCOPUS database on October 16, 2021, within the article title (TITLE), abstract (ABS), and keywords (KEY): “consumer engagement”, OR “consumer brand engagement”, OR “customer brand engagement”, OR “customer engagement behavior”/OR “behaviour”, OR “consumer engagement behavior”/OR “behaviour”, OR “brand community engagement”, specifying the document type only articles and the time period from the first article in the database until the entire period (except 2022), we have found 2201 articles. We removed 5 duplicates and 18 non-English papers during the screening phase and at the third, eligibility phase we cleared the database from 38 editorials, book reviews, and systematic literature reviews. According to Lim & Rasul (2022), the eligibility phase is necessary so that there is no repetition in the results. Lastly, to decide which articles to include in the further analysis we applied content analysis to the remaining abstracts and thus removed 220 abstracts as inadequate for analysis. The final data set is comprised of 1920 journal articles.

The data descriptive analytics is performed in Excel. The country co-authorship and keywords co-occurrence maps are created in the software VOSviewer version 1.6.18. Details for the software can be found in Van Eck and Waltman (2010). The machine learning algorithm for logistic regression to find the key influencers for citation is performed in the software tool Power BI Desktop.
4. RESULTS AND ANALYSIS

4.1. Descriptive analytics
In the analyzed period (2006-2021) 1920 articles were published. The highest number of published articles is in the last year (410 articles). In addition, in the last five years (2017-2021) 77.03% (1479) of the articles are published implying a high research interest in this area.

![Figure 1: Distribution of articles by year](image)

We have identified the top five journals based on the number of published articles in the analyzed area (Figure 2). In those journals, 242 articles were published representing 13% of the total number of published articles. Based on the artificial intelligence (AI) driven visual decomposition tree in Figure 2, we can see that the Journal of Business Research is a leader with 69 articles (29% of the total number of articles in the top five journals), followed by the Journal of Retailing and Consumer Services (55 articles, i.e., 22.7%), the Journal of Product and Brand Management (50 articles, i.e., 20.7%), the Journal of Research in Interactive Marketing (35 articles, i.e., 14.5%), and the Journal of Services Marketing (33 articles, i.e., 13.6%). Based on our findings, 2021 is the best year in terms of the number of published articles for the Journal of Business Research, the Journal of Retailing and Consumer Services, and the Journal of Research in Interactive Marketing (61 in total).

![Figure 2: Top five journals based on the number of published articles](image)
The ten most cited articles are shown in Table 1. The most cited article (1533 citations) is “Customer engagement behavior: Theoretical foundations and research directions” by Van Doorn et al. (2010) published in the Journal of Service Research. The same journal also published the second most cited article (1449 citations) written by Brodie et al. (2011).

<table>
<thead>
<tr>
<th>Title</th>
<th>Source title</th>
<th>Authors</th>
<th>Year</th>
<th>Cited by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer brand engagement in social media: Conceptualization, scale development and validation</td>
<td>Journal of Interactive Marketing</td>
<td>Hollebeek L.D., Glynn M.S., Brodie R.J.</td>
<td>2014</td>
<td>1035</td>
</tr>
<tr>
<td>We’re all connected: The power of the social media ecosystem</td>
<td>Business Horizons</td>
<td>Hanna R., Rohm A., Crittenden V.L.</td>
<td>2011</td>
<td>806</td>
</tr>
<tr>
<td>Undervalued or overvalued customers: Capturing total customer engagement value</td>
<td>Journal of Service Research</td>
<td>Kumar V., Aksoy L., Donkers B., Venkatesan R., Wiesel T., Tillmanns S.</td>
<td>2010</td>
<td>631</td>
</tr>
<tr>
<td>The process of customer engagement: A conceptual framework</td>
<td>Journal of Marketing Theory and Practice</td>
<td>Bowden J.</td>
<td>2009</td>
<td>601</td>
</tr>
<tr>
<td>Customer engagement, buyer-seller relationships, and social media</td>
<td>Management Decision</td>
<td>Sashi C.M.</td>
<td>2012</td>
<td>583</td>
</tr>
<tr>
<td>A descriptive model of the consumer co-production process</td>
<td>Journal of the Academy of Marketing Science</td>
<td>Etgar M.</td>
<td>2008</td>
<td>551</td>
</tr>
</tbody>
</table>

We have created a country co-authorship network visualization map by using the bibliographic data, setting the type of analysis to be "co-authorship," the units of analysis to be "countries," the counting method to be "full counting," the maximum number of countries per document to be "5", and the minimum number of documents and citations for a country to be "1". There have been 104 terms identified, but we have focused on the top ten countries based on the number of published articles that comprise the co-authorship network visualization map presented in Figure 3. The countries are grouped into 3 clusters so that the countries with the same color are part of the same cluster. Cluster 1 (the red cluster) includes Australia, Canada, China, Malaysia, and the United States. Three countries belong to cluster 2 (the green cluster): Portugal, Spain, and the United Kingdom, while two countries belong to cluster 3 (the blue cluster): France and India. Additionally, in Table 2, we present the number of articles for each of those 10 countries (blue bars), as well as the top 10 countries with the highest number of citations (orange bars). Based on Figure 3, we can see that the largest item is labeled as "United States" and the weight of the item is linked with the number of articles, i.e., as presented in Table 2, the United States has 514 published articles but also is the leader regarding the citations (20478). Australia, which also belongs in the red cluster, is in second place based on the number of published articles (263) and citations (7056). The United Kingdom, which belongs to the green cluster, is in third place based on the number of documents (227) as well as the number of citations (6669). Regarding the number of documents, the fourth-ranked country belongs in the blue cluster, that is India (198), but it is seventh-ranked based on citations (3291), etc. Also, based on Figure 3, we can see that the United States has the highest link strength with Australia (link strength of 29), while their
link strength with Canada is 11. In addition, the link strength between France and India is 13, while that between France and Spain is 2. Based on this, we can state that the short distance between countries is not always crucial for co-authorship.

*Figure 3: Country co-authorship network visualization map*

4.2. Keywords co-occurrence analysis based on text mining in the abstracts

In this section we present two network visualization maps regarding the keyword co-occurrence for the entire analyzed period (2006-2021) and the last three years (2019-2021).

4.2.1. Keywords co-occurrence analysis based on text mining in the abstracts for the whole analyzed period (2006-2021)

The map is created by mining the text data, i.e abstract data, so, the structured abstract labels and copyright statements have been ignored. The counting method is binary. From the 1920 mined abstracts 30563 keywords are extracted, and the minimum number of occurrences of a keyword is set at 10, so that 951 keywords meet the threshold. For each of these 951 keywords a relevance score is calculated, and the software selects the 60% most relevant ones based on the relevance score, so that 571 keywords remain. After a detailed analysis of the 571 keywords, we have excluded 133 which are not relevant for the analysis (the complete list can be provided on request). The keywords co-occurrence network visualization map consists of 438 keywords that are grouped in four clusters.

Table 2: Countries with the highest number of articles and citations

<table>
<thead>
<tr>
<th>Country</th>
<th>Articles</th>
<th>Country</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>514</td>
<td>United States</td>
<td>20478</td>
</tr>
<tr>
<td>Australia</td>
<td>263</td>
<td>Australia</td>
<td>7056</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>227</td>
<td>United Kingdom</td>
<td>6669</td>
</tr>
<tr>
<td>India</td>
<td>198</td>
<td>New Zealand</td>
<td>6234</td>
</tr>
<tr>
<td>China</td>
<td>156</td>
<td>Germany</td>
<td>5353</td>
</tr>
<tr>
<td>Spain</td>
<td>81</td>
<td>Netherlands</td>
<td>4820</td>
</tr>
<tr>
<td>Malaysia</td>
<td>73</td>
<td>India</td>
<td>3291</td>
</tr>
<tr>
<td>Canada</td>
<td>64</td>
<td>China</td>
<td>3010</td>
</tr>
<tr>
<td>France</td>
<td>61</td>
<td>France</td>
<td>2334</td>
</tr>
<tr>
<td>Portugal</td>
<td>57</td>
<td>Finland</td>
<td>2228</td>
</tr>
</tbody>
</table>
Each item on the network visualization map is presented with a label and a circle. The size of the labels and circle depends on the weight of the item. The higher the weight (that is linked with the occurrences of the item) the larger the circle and its label. In addition, two items are linked with a line between them, so that the stronger the link the shorter the distance between the items. If the label is not presented, that is to avoid the problem of overlapping labels.

Based on Figure 4, we can see that the item, i.e., the keyword “brand engagement” that is part of the red cluster is the largest one on the whole map, while “Facebook” is the largest one in the yellow cluster, “interview” in the blue cluster and “technology” in the green cluster.

Figure 4: Keyword co-occurrence network visualization map for the period 2006-2021

The keyword co-occurrence in the red cluster indicates that CE studies in this cluster examine how “brand engagement” on “social media platforms” can influence “service quality”, “customer satisfaction” and “customer loyalty”. In this cluster, the “questionnaire” is the most applied data collection form, while “confirmatory factor analysis”, “PLS” and “SEM” were mostly used for data analysis. The most analyzed context in this cluster is the “hotel” industry. The studies in the yellow cluster focused on explaining how the “branded content” format (“video”, “picture”, “photo”) is related to “consumer response” and “reaction” on “message”/“post” on “Facebook” and “Instagram”. The keyword co-occurrence in the blue cluster indicates that studies in this cluster investigate how “recommendation” influences “group” “assessment” and “belief”, focusing on different “age” groups (“young person” and “adults”), mostly collecting data through “interview”. Also, it is evident that the main industry analyzed in this cluster is “health service” and “care” where the research focus is on developing “program” for “training” the “staff” and “health professionals” as a “policy” for building “partnership” and thus improving engagement. The green cluster comprises CE studies that are related to “technology” and “innovation” as means of building “capacity” and improving “system” “efficiency”. “Energy” is the most analyzed sector in this cluster.

4.2.2. Keywords co-occurrence analysis based on text mining in the last three years (2019-2021)

In the last three years 1111 articles were published, and we have used their abstracts to perform text mining and identify research hotspots in the most recent period. We have used the same procedure as in mining the abstracts for the whole period, except the threshold for occurrences is set at minimum five. So, in total, 19408 items were extracted, and minimum
1187 items occur 5 times. Of these 1187 items, 712 are the 60% most relevant ones based on the relevance score. After a detailed analysis, we have excluded 100 items that are not relevant for the analysis (the complete list can be provided on request). The keywords co-occurrence network visualization map consists of 612 keywords. The keywords are grouped into 6 clusters, with “interview”, “Facebook”, “brand loyalty”, “system”, “customer brand engagement”, and “challenge” being the largest keywords in the green, violet, red, light blue, yellow and blue clusters respectively. In addition, in the last year, the researchers focused mostly on the pandemic, and in 2020 they used artificial intelligence (AI), chatbot and structural equation modeling (SEM).

**Figure 5: Keyword co-occurrence network visualization map for the last three years**

4.2.3. What influence a citation on CE to increase?  
We have used the AI driven visual in the Power BI Desktop, known as a decomposition tree to explore the distribution of the data, where we have analyzed the citations regarding the publisher. The publisher with the highest no. of citations is Elsevier Ltd. (6375 citations).

**Figure 6: Key influencers for citations**

In addition, to answer the third research question in our study, we have performed a machine learning algorithm for logistic regression in Power BI Desktop, where we have used the number of citations of each paper as a dependent variable, and the following variables: number of authors per paper, authors affiliations, publisher, source title and year as
independent variables. Our question in the AI driven tool Key influencers was what influences citations to increase? Based on the performed logistic regression (the AI visual Key influencers), we can see that when the year is 2016 or less, the average of citations increases by 45.51x (Figure 6).

5. CONCLUSION
This study provides a bibliometric analysis of CE articles in English by searching for CE-related terms in the SCOPUS database. The analyzed period is 2006–2021, and the sample for analysis after the incorporation of the PRISMA protocol is 1920. It presents the first CE bibliometric study with the largest sample included, and despite the descriptive analysis, county co-authorship, and keyword co-occurrence maps based on text mining, it uses a machine learning algorithm for logistic regression to analyze what influences the citations in this area to increase.

Our findings show: 1) that the researchers' interest in this area has boomed in the last five years when they published 77.03% of the articles, with 2021 being on the pedestal as a year when 410 articles were published; 2) the top five journals based on the number of published articles, so that the leader is the Journal of Business Research with 69 articles; and 3) the top ten most cited articles so that the first ranked is the article written by Van Doorn et al. (2010) with 1533 citations. Those articles should be used by researchers interested in CE research as a starting point in conducting research. 4) The country with the highest number of CE-related articles (514) is the United States, and it also has the highest number of citations (20478); 5) the keyword co-occurrence network visualization for the entire analyzed period is comprised of four clusters, with the most used term: "brand engagement", which belongs to one cluster, and the three largest terms (each in one cluster) are: "Facebook", "interview" and "technology". 5) The keywords co-occurrence network visualization map for the period 2019–2021 consists of 6 clusters, in 2020 AI, chatbot, and structural equation modeling are used as methods for analysis, and in 2021 the research is mostly about pandemic; 6) the AI-driven visual, decomposition tree shows that Elsevier Ltd. is the publisher with the highest number of citations (6375); 7) the year, i.e., 2016 or less, is identified as a key influencer of the citation increase based on the machine learning algorithm for logistic regression.

In this study, we only focus on articles in English that are published in journals and indexed in the SCOPUS database. Therefore, we might have omitted articles in other languages and indexed in other databases such as Google Scholar, etc., which we provide as a limitation of the conducted study. Also, future studies should focus on identifying the main theories in CE studies, as well as the antecedents and consequences of this concept. Considering that the era of automation is coming rapidly, future research should consider how AI helps companies increase customer engagement.

REFERENCES


