Social Media CSR Communications: Comparative Content Analysis via Topic Modellingacross Four Industries

Rebecca Lim

ABSTRACT

This paper uses topic modelling to observe the content type distribution across companies' CSR communications on the platform Twitter, focusing on four main industries : Entertainment, Beauty, Retail, and tech. Industry groupings as well as account type groupings are used to model topics that span across three main dimensions of CSR within the triple bottom line model: environmental, social, and economic sustainability. Findings reflect the strength of CSR-dedicated accounts in expressing salient and focused topics, while showing room for representation of economic sustainability within CSR communications. General accounts seem to be focused on brand promotion over CSR related content, but depending on industry target demographics, sustainability content representation can be more prominent.

KEYWORDS

Sustainability, LDA, Topic, Business Communication, Twitter, Sustainable Computing

INTRODUCTION

Rising concerns regarding the detrimental effects of modern environmental degradation have led to increasing public and legislative demands for organizations to incorporate sustainability in their activities (Freundlieb 2013). Modern consumers have shown to take into consideration the extent of enterprises' commitments to social anl and environmental responsibilities; as a result, many firms have acknowledged that their response to sustainability demands have the potential to influence their competitiveness, firm reputation, and even financial performances(Koller 2011; Agudelo 2019). Such demands and commitments are often referred to as a process of Corporate Social Responsibility-- as commonly defined according to some of the earliest business and sustainability principles, CSR is the pursuit of achieving “triple bottom line” in business activities, incorporating social, environmental, and economic sustainability (Elkington 1998). The extent to which these three aspects are highlighted or equalized depends on the defining authors, and similarly, emphasis and modes of accomplishment within sustainability initiatives vary across enterprises and industry. Further, the final communication of these sustainability commitments is crucial to stakeholders' reception of the initiative; a lacking method and quality of communication may lead to poor public and consumer perception, especially when signalling insincerity or poor follow-through(Porter and Kramer 2006).

With the increase in web-based platforms and communications, firms have taken advantage of tools such as social media to implement their CSR initiatives. Social media channels provide specific benefits for the communication of a firm's effort towards sustainability, such as high volume contact and interaction, real-time and high frequency flow of communication, incorporation of voice and brand identity, and access to engagement metrics and data(Etter et al. 2018). Social media specifically extends the audience of information to web-based interpersonal connections, uprooting prior models of messaging to well-acquainted consumers and providing the possibility for a much wider online network of interconnected social groups for information transmission(Freeman and Moutchnik 2013). The firm-to-stakeholder contact social media platforms is also utilized by organizations and firms to give voice to consumers and their needs regarding CSR, although the stakeholders themselves may not in reality be elevated to the actual decision making processes (Fuchs 2009). The integration of this tool into business sustainability does promise unique benefits, but the exact scale and accessibility of this medium also requires firms to strategize carefully about successful communication of firm information, especially that of which significantly affects consumer appeal(Lee et al. 2013).

Extant literature has documented the effects and features of both sustainability reports and online CSR communications, exploring the various qualities of either channel that enable successful engagement, messaging, and brand legitimization. Reports have been analyzed to reflect their language-based sentiments, effects of assurance statuses, and quality of information based on varied itemized scales((Saeli 2019;Silvia Romero et al. 2014 ;Freundlieb and Teuteberg 2012). CSR Social Media, on the other hand, has been analyzed in various mediums(Twitter, Facebook, etc.) in terms of the incorporated voice/identity as well as firms' level of new media command over sustainability communications(Flora Hung-Baesecke et al. 2016; Saxton et al. 2019; Chae and Park 2018). However, there is limited knowledge on the range of contents selected by firms to present on these platforms, specifically by CSR dimension or cause. While communications such as static reports necessitate a complete overview of a firm's CSR activities over a given time frame, social media updates vary in their coverage based on the choice of companies and their representatives. The type of CSR account also varies by company, ranging from general CSR accounts to ; other companies choose to report CSR content on their general social media accounts, choosing not to designate platforms for sustainability updates altogether. Firms exercise choice in both content topic range and account type as a part of CSR strategy, and the results ultimately shape the online representation of a company's overall CSR efforts. Dyllick and Hockerts' framework asserts that the three dimensions of CSR, although distinct and separate in an operational lens, ought to be present and integrated in a strategic approach to sustainability(Dyllick and Hockerts 2002). Building upon this framework, I seek to explore which dimensions are prominent in the contents which are posted on social media communications. Which social, environmental, and economic aspects of sustainability are being communicated by firms? Which topics do firms focus on communicating via social media, and what can trends in the content illuminate about their messaging objectives?

With the research and discussion above, I pose a central research question and two subquestions:

CRQ1: What topic contents are companies selecting to address their range of CSR initiatives on their social media platforms?

SQ1: What are the topic distributions within social media CSR communication contents of four leading industries, and how do they compare?

SQ2: Are there significant differences between the topic content between main corporate and CSR dedicated accounts?

As the main objective of the study is to observe the topic distributions in reported and communicated sustainability topics between two mediums, the broader implication of the findings will be to provide knowledge regarding the prioritization of sustainability themes within CSR communications, which can then be applied to improve online sustainability initiative messaging. Through the impact assessment according to Stakeholder Theory, I assert that the success of these communications will contribute to a key imperative of CSR, which is to benefit both stakeholder and firm success through positive feedback of expressed needs and the accomplishment of those needs(Grau 1970). Analysis of the differentiation between topic distributions is likely to show where social media content is most sparse within the industry or companies' communications, in terms of the breadth of reported topics. Identification of these gaps in content material can help develop suggestions for future CSR communications, especially regarding firms' CSR efforts that are conducted but poorly communicated. Prior studies indicate ways in which the development of social media CSR communications can be strengthened. I build upon the few topic analyses in the existing literature to modify our methods; while prior studies have documented the topic trends for decades at a time and across a general and broad catalog of companies, this study aims to focus on the most recent years of social media content prominence among select leading industries and the representative companies which exhibit social media CSR communications to achieve a holistic enterprise sustainability commitment(Székely and vom Brocke 2017). Overall, this study is significant as a focused comparison of CSR topic coverage among digitally active, and depending on the results, I seek to provide useful suggestions for online sustainability communication coverage and anticipation for the future trends of CSR web content.

METHODS

Preparation & Initial Sampling

In order to sample companies for the study, I compiled the leading sustainability and CSR ranking organizations' lists from Forbes and Newsweek and selected four prominent industries of interest : Technology(Tech), Beauty, Entertainment and Retail.(Todd 2020; Newsweek 2017). For the selection of companies, the lists were observed following the ranking order, followed by a screening for eligibility: to be included in the sample, each company needed to have an active corporate or CSR/Sustainability Twitter account that has generated content in the last 3 years. To supplement for eligible companies, alternative listings were consulted specific to the industry for prominent CSR activity and branding. After screening, I collected 3-4 main accounts per industry category(Variety, 2016). The selected industries and companies and their accounts are as follows:

Sample Companies and Accounts per Industry

Industry	Technology	Beauty	Entertainment	Retail
Companies & Accounts	1.IBM(@IBM) 2. Dell (@DellTech) 3. Microsoft (@Microsoft_Green)	1. Love, Beauty, and Planet (@beautyand planet) 2. L'oreal (@lorealcommitted) 3. Mac (@maccosmetics)	1. Disney (@DisneyCSR, @DisneyConserves) 2.Warner Brothers (@warnerbros) 3. Lego (@legofoundation)	1. Adidas (@adidas) 2) The North Face (@thenorthface) 3) Patagonia (@patagonia) 4) Reformation (@reformation)

Among potential platforms, Twitter was chosen as the social media platform due to its prominence as an enterprise communication method compared to other alternatives, as well as the accessibility of the Application Programming Interface (API) and other alternative forms of data public data scraping tools developed for the site that has historically been used for similar types of research.

Data Collection

To collect the necessary data for topic modelling and analysis, I used Twitter.com to locate each account. For this collection step, the account names were recorded to input into the Tweepy models, which are built using publicly available Tweepy python library packages designed to access Twitter API. To retrieve the tweets uploaded within the desired timeframe of the past 3 years, the variables "startDate" and "endDate" in the Tweepy model were set to values corresponding to May 2017 and May 2020 respectively.

Pre-Processing

After inputting the desired timeframe and account names, the Tweepy program was run and tweets were retrieved in raw text format. This text needed cleaning before it could be processed to create topic models, as it contained many words and characters irrelevant to content analysis. I cleaned the resulting data for excess and filler language using implementations in Gensim as well as publically available Python scripts developed by other students and researchers(Github repositories attributed to @hadiz and @ariddell). The full Github repositories and links can be found within the resources section, for which all authors and teams are responsible. These programs are built upon existing guides and topic modelling tutorials to fit social media analysis needs. (Saxton 2018; Li 2018).

For tweets, data from the geTweets process was stored for each tweet, and python script was used to extract the files into a standard text document. Words were normalized around the intended terms, documents were also cleaned for any non-English language, and stop words were removed. The cleaning specific to tweets include omitting factors such as usernames, numbers, and special symbols/characters were filtered out. Previous research has recognized that the unique format of Tweets, which are often abridged or shortened messages, pose the problem of topic sparsity in large-scale topic analysis(Bicalho et al. 2017). The end product was a fully pre-processed and cleaned Master Tweets Corpus.

Topic Modelling

With the pre-processed data, I utilized the aforementioned publicly available python scripts in order to accomplish the Latent Dirichlet Association (LDA) topic modelling and retrieve results, then manually assigned topics. The LDA method of topic modelling is used to identify prevalent themes and topics within a collection of documents, also used to gauge meaning and sentiment in natural language datasets(Saxton 2018). With the LDA modelling package retrieved, I entered the corpuses through the main algorithm. In order to address my two subquestions, I chose three groups of corpuses to build models around: first, industry groupings, second, companies by account type, and third, individual companies. Industry groupings were included so I could model around the topics most prominent across all the given sector/industry's companies. The account type groupings allowed for separate modelling of corporate accounts and dedicated CSR accounts. Individual company corpuses allowed for modelling for singular account data, which could be observed to point out anomalies or special clusters within that could have influenced the former models. For all documents, based on the coherence values per topic number, I chose the most appropriate standard dimensions, which have been identified as when the point at which coherence values maximize, then begin to dip. The final model produces a topic probability distribution over each document, as well as a list of unnamed topics identified by the most probable associated words, which are also represented by the percentage likelihood of appearance. Based on the most probable associated words per unnamed topic, I named each grouping with the most relevant and appropriate label, consulting the GRI indicator standards as well as online dictionaries in greyer areas of interpretation. Next, I clustered these topics based on similarity and association (i.e. Philanthropy, Workers' rights, etc.), making note of environmental/ecological/social categorizations based on GRI indicator guidelines, which helped clarify more specific trends in selective CSR topics per grouping. At the end of this step, each document has its most prominent topics and corresponding categories named and identified.

Final Visualization and Analysis

pyLDAvis Visualization

To visualize the results, the package pyLDAvis was utilized for a rudimentary representation of the topic distribution per document. As LDA models are processed through pyLDAvis, the result is an interactive two-part visualization which includes an intertopic distance map with topics represented as circles scaled to their frequency of occurrence, as well as a graph showing the most relevant terms which generate each topic. In theory, the pyLDAvis is adequate for the most fundamental topic distribution data, but because it is not a static result, it is difficult to transfer as a figure. However, the intertopic distance models were paired with the percentage values of topic prominence to solve this issue. Charts were created to show quantitative percentage values of how much of each given document pertained to the top topics, showing the prominence of each topic as well as corresponding key words. Topic codes corresponded to those of the pyLDAvis results. The pyLDAvis was therefore conducted with models for all three groupings: industry, account type, and individual company. The data for individual companies and account type groupings were analyzed via pyLDAvis only, while the industry data was used to create extra visualizations in order to analyze trends for both subquestions.

Word Distribution Visualizations

Two types of additional visualizations were produced consulting online modelling visualization guides provided by the educational site, machinelearningplus.com, with variables manipulated to fit the existing script. Both supplement the pyLDAvis for a closer understanding of topic constituents as well as prominence over groupings.

First, graphs of word distributions were created to detail the contents of each topic per document across groupings. The resulting figure organized the word frequency and weight for all key words within a given topic, also following previously generated topic codes. Upon generating the topics, I approached some issues in terms of the visualizations for the account type grouping, due to the truncated data. However, I was able to analyze based on the original LDA model data by manually retrieving prominent topic data without the aid of visualizations.

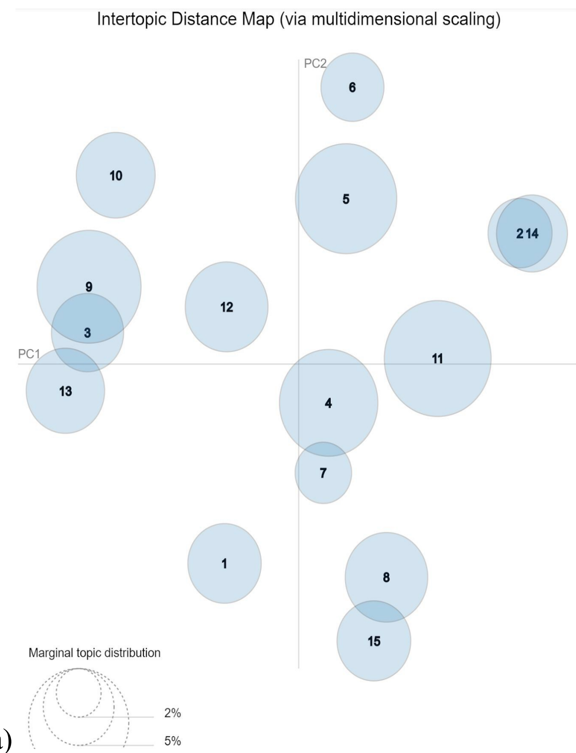
Second, as a qualitative visual representation of the most salient topics, word clouds were generated for the top 10 topics per document using the online modelling visualization guides. Text colors were manipulated from existing code for aesthetic purposes.

RESULTS

pyLDAvis Visualizations

Industry pyLDAvis results showed two key types of topic distributions. The Tech and Beauty Industry intertopic maps showed a diverse spread of topic prominence, such that the circles in the distribution showed similar sizes with no clear sign of dominance from one topic alone(Figures 1(a)-1(b)). Meanwhile, the Entertainment and Retail Industry intertopic maps showed a clear dominance of a single topic, which appeared in a frequency of nearly 100%, throughout the document, as indicated by the topic circle (Figure 1(c)-1(d)). The first type of topic distribution reflects that the document contents differentiate from each other enough in meaningful clusters, without one topic capturing the major sentiment of all tokens. Distance between the topics were also significantly greater in Figure 1(c) and Figure 1(d), showing that all other topics showed little to no likeness to dominant topic contents.

Further, the chart visualization results accompanying the intertopic distance map captured all 15 selected topics for Figures 1(a)-1(b), while only the dominant topic was represented for 1(c)-1(d). For the latter group, the fraction of dominant topic occurrence within the document was close to 1.0, omitting all other topic which exhibited an insignificant occurrence frequency throughout.



a)

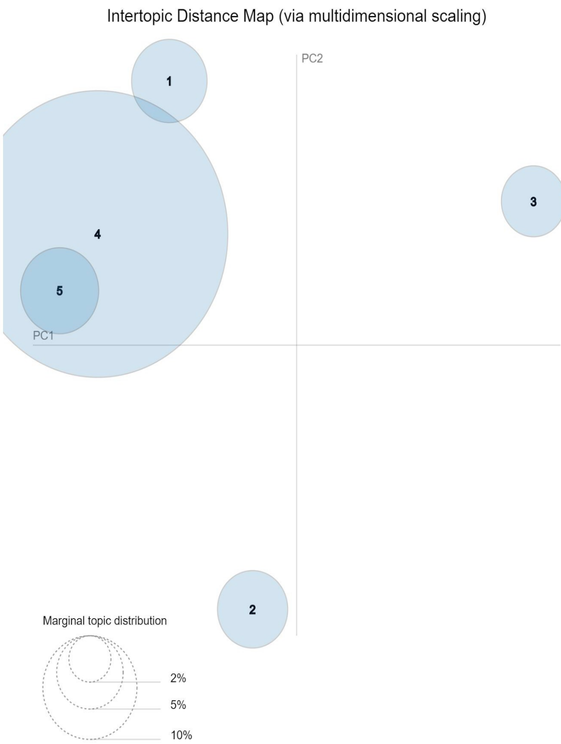
Dominant_Topic	Fraction_of_Documents	Topic_Keywords
0	0.058043	read, drug, scientist, learn, target, potentia...
1	0.050417	year, support, celebrate, award, proud, join, ...
2	0.024714	learn, question, answer, insulin, program, cal...
3	0.077673	today, science, event, partner, start, year, h...
4	0.125830	research, patient, treatment, cancer, advance,...
5	0.061291	healthcare, innovation, health, technology, ca...
6	0.081627	live, today, share, day, patient, support, wat...
7	0.061291	jnj, tip, discover, learn, product, make, baby...
8	0.071176	woman, learn, work, scientist, nurse, meet, di...
9	0.060585	health, learn, people, world, work, programme,...
10	0.062138	employee, work, family, learn, day, share, com...
11	0.044344	disease, people, month, condition, affect, lea...
12	0.055218	company, health, top, proud, sustainability, l...
13	0.072165	patient, disease, risk, address, improve, asth...
14	0.058890	cancer, patient, treatment, cell, lungcancer, ...
15	0.034600	learn, vaccine, covid, watch, global, support,...

Figure 1(a). Topic prominence and most salient terms per topic for Tech Industry. Intertopic distance map generates axes based on similarity of topics to each other based on shared keywords and prominent sentences, and size of the topic circles indicate approximate percentages of the document that pertain to each topic. Chart matching each map shows the quantitative percentage values for topic prominence in each document, along with main key words which constitute the topic. Topic count is set to 15 for Tech and map shows diverse spread across intertopic distance. Results share identical form for remaining industries in figures 1(b)-1(d).

	Fraction_of_Documents	Topic_Keywords
Dominant_Topic		
0	0.091944	fragrance, shop, scent, rise, hair, collection...
1	0.158583	oreal, woman, world, support, company, commitm...
2	0.174188	shop, lip, color, shade, envy, love, lipstick,...
3	0.079291	double_wear, foundation, find, shade, flawless...
4	0.136229	skin, shop, repair, night, serum, glow, day, e...
5	0.120202	shop, beauty, makeup, collection, gift, estee_...
6	0.073809	touch, purchase, learn, experience, find, skin...
7	0.105441	oreal, product, reduce, learn, sustainable, pa...
8	0.060312	smallactsoflove, make, planet, love, day, recy...

b)

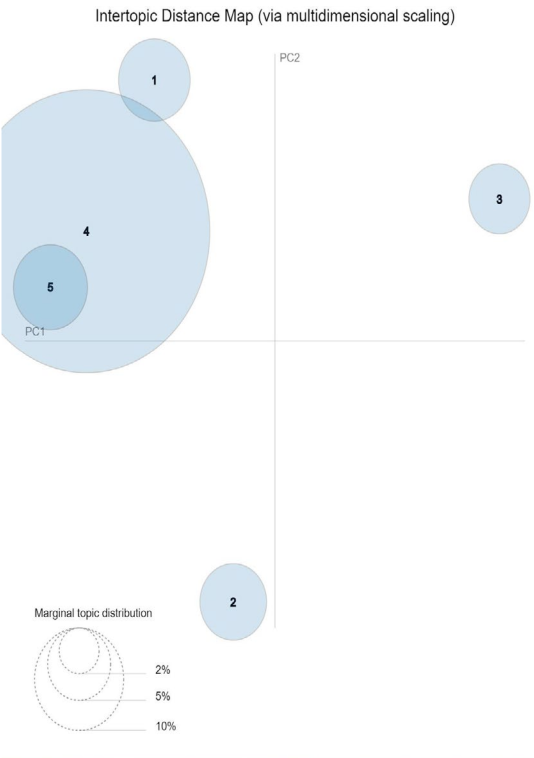
Figure 1(b). Topic prominence and most salient terms per topic for Beauty Industry. Intertopic map shows 15 topic with more polar spread in two topic similar areas, whie chart shows dominance for 8 topics, with the rest omitted for low frequency.



c)

	Fraction_of_Documents	Topic_Keywords
Dominant_Topic		
3	1.0	play, learn, child, learning, education, work,...

Figure 1(c). Topic prominence and most salient terms per topic for Entertainment Industry. Map and chart show the dominance of topic 4, with nearly ~ 100% dominance within the document. One subtopic exists within the dominant topic, and others share little similarity as shown by distance within the map.

d) 

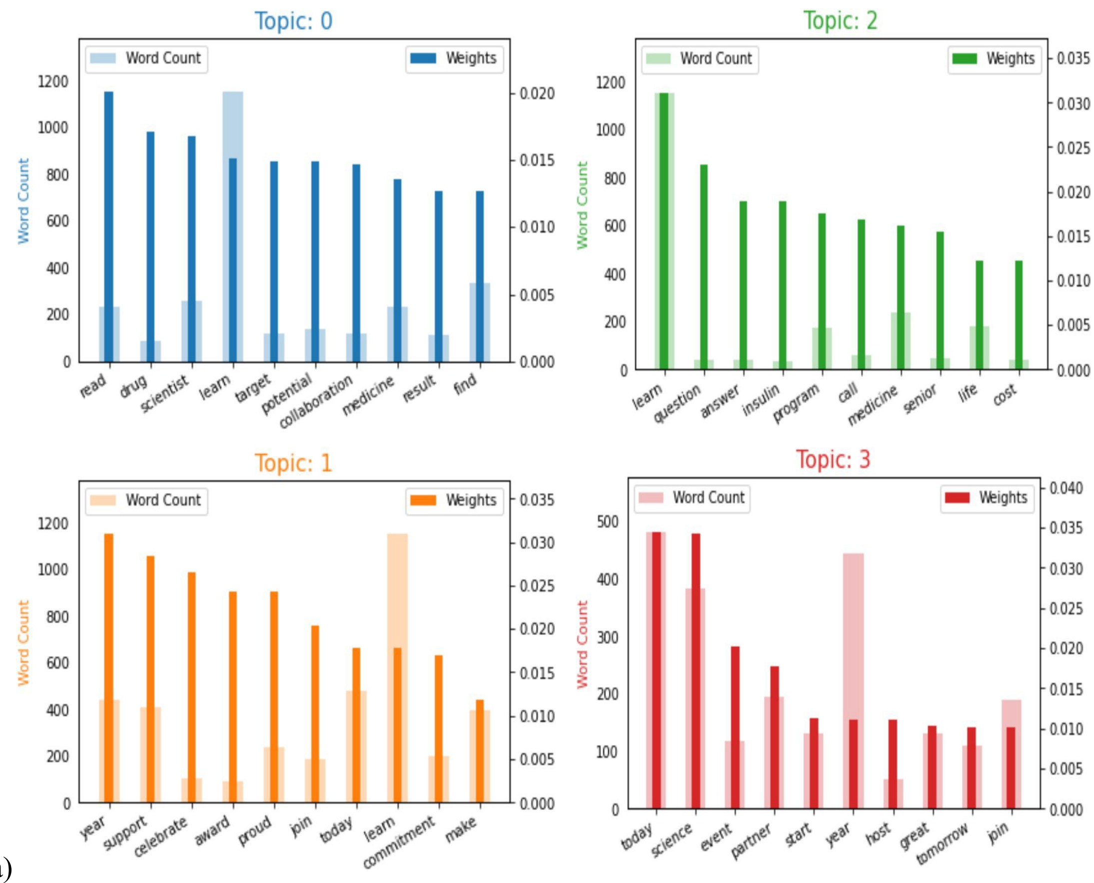
Dominant_Topic	Fraction_of_Documents	Topic_Keywords
4	1.0	photo, make, collection, create, good, time, y...

Figure 1(d). Topic prominence and most salient terms per topic for Retail Industry. Similarly to figure 1(a), a single topic exhibits full dominance.

Word Distribution Visualizations

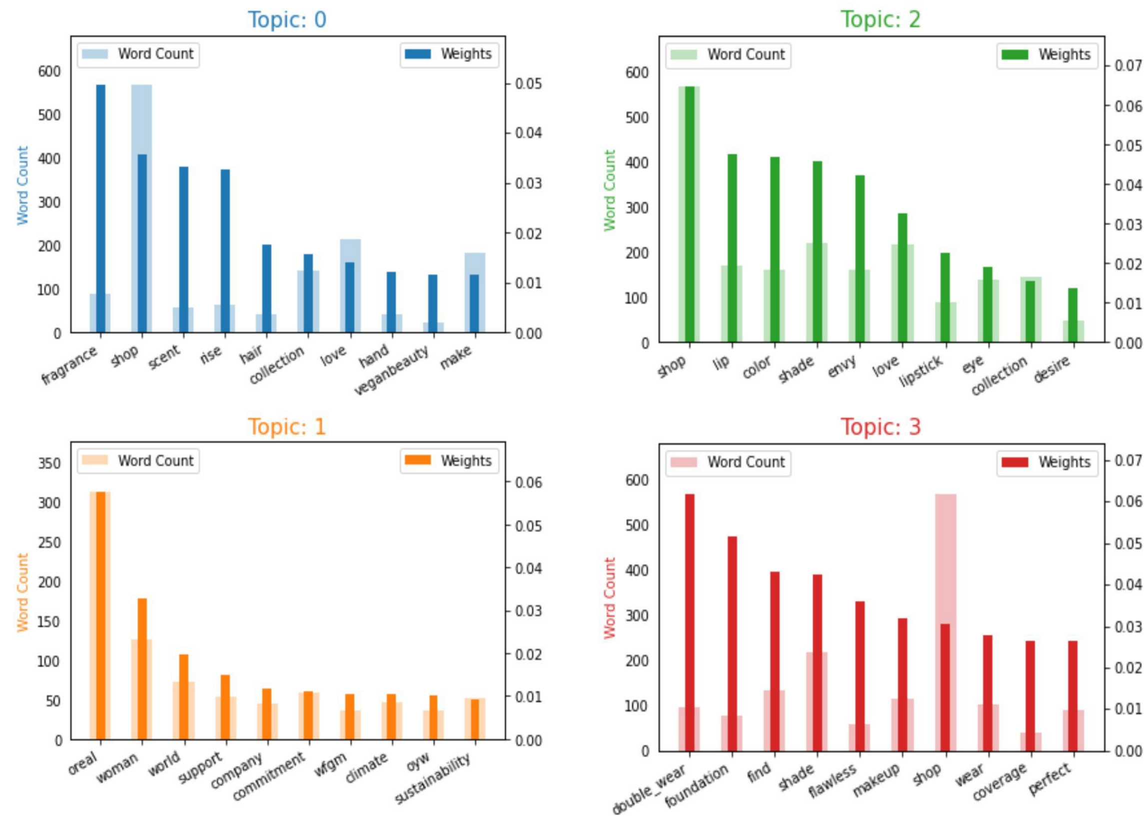
First, word distribution per topic visualizations showed topic saliency within the top 4 topics per industry. Visualizations for the Tech, Entertainment, and Beauty Industry groupings exhibit the most salient topics, with one word showing highest frequency and the following ranked topic showing a curve shaped decline(Figure 2(a), 2(b), 2(c)). The visualization for the Retail Industry grouping however, shows word weight and frequency that varies somewhat randomly among the top words per topic. While the words are still ranked based on frequency, the bar chart shows a flatter shape with a few peaks, showing that word frequency is similar for the top cluster.

Second, word clouds show the most salient topic for each industry grouping. The Tech Industry grouping cloud shows a topic related to women and girls in stem fields and careers, with a spread of gendered and industry related terms(Figure 3(a)). The Beauty Industry grouping cloud shows the most salient topic involving terms related to individual activism and contribution to environmental cause(Figure 3(b)). The Entertainment Industry grouping cloud shows a topic peripheral to the Tech Industry cloud topic in terms of dimension, with terms relating to childrens' programs and education(Figure 3(c)). Lastly, the Retail Industry grouping cloud shows a topic with key terms relating to political activism and voter rights, the first to include a political dimension(Figure 3(d)).



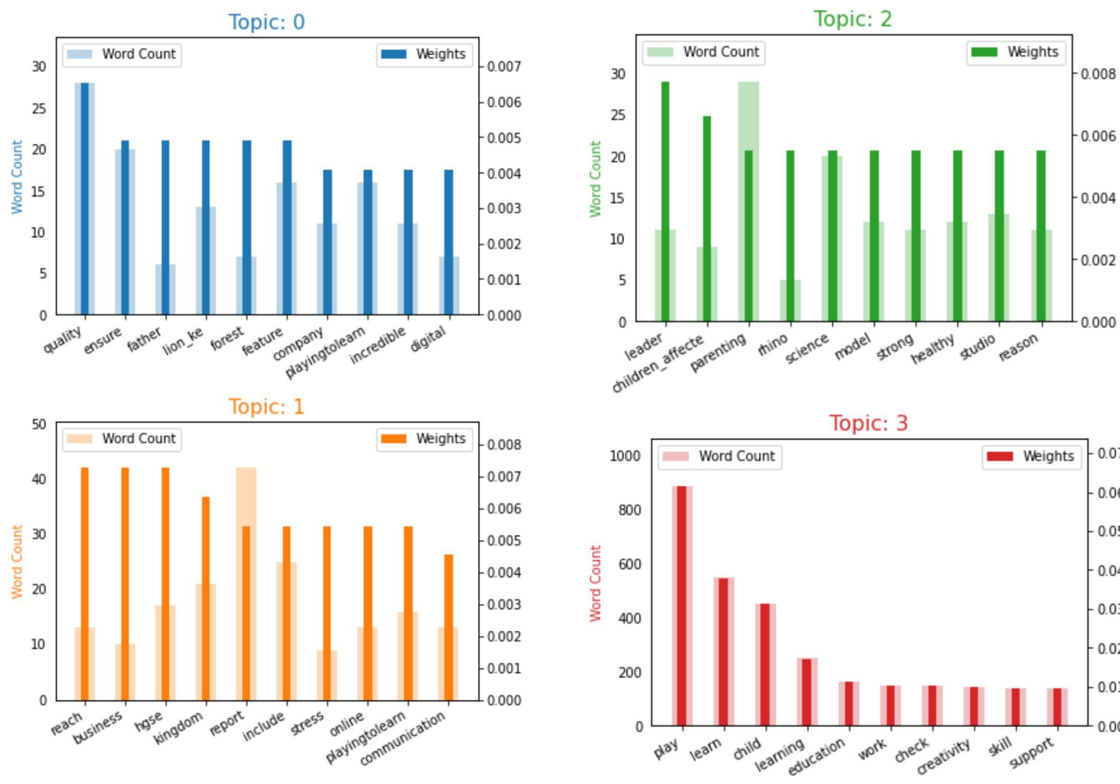
a)

Figure 2(a). **Top topics per document and corresponding weight and frequency for Tech Industry.** Per each industry grouping, graphs show top 4 topics and the frequency in word appearance and weight as measure in decimal value for contribution to the entire topic catalog of key words. Topics 0-4 show somewhat even spread of word frequency.



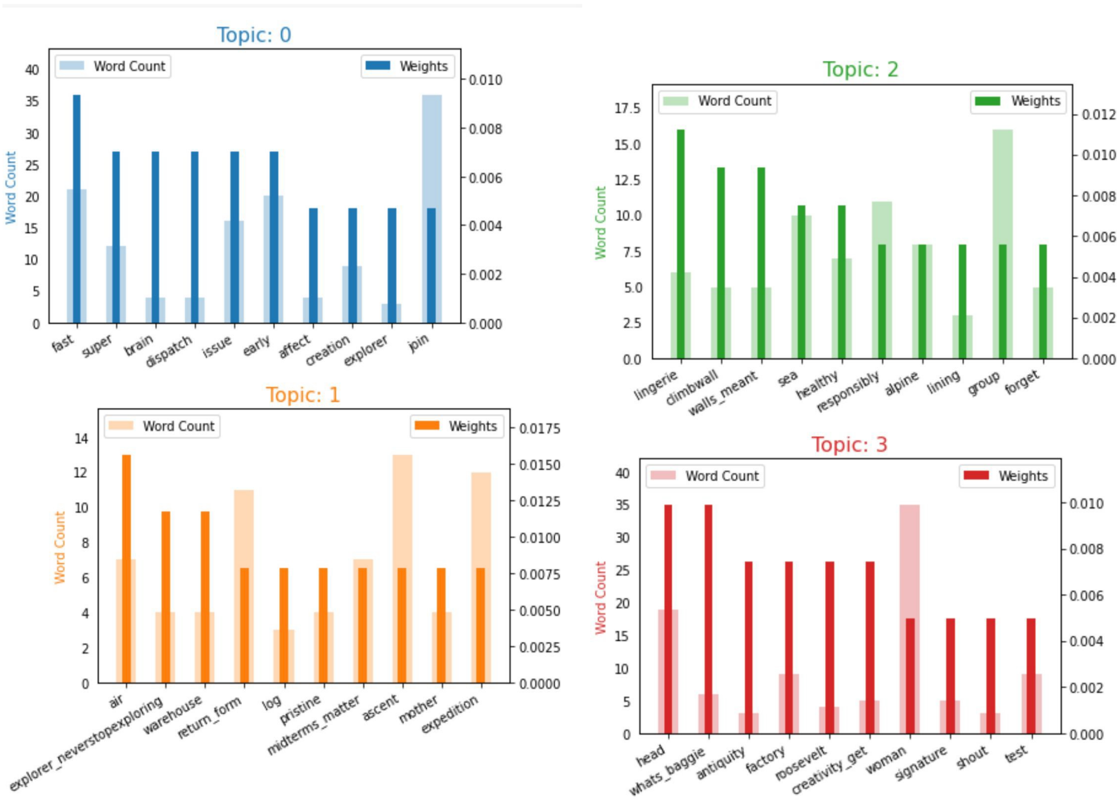
2(b)

Figure 2(b). **Top topics per document and corresponding weight and frequency for Entertainment Industry.** Topics 1 and 2 show topics with singular dominant word, with falling frequency thereafter.



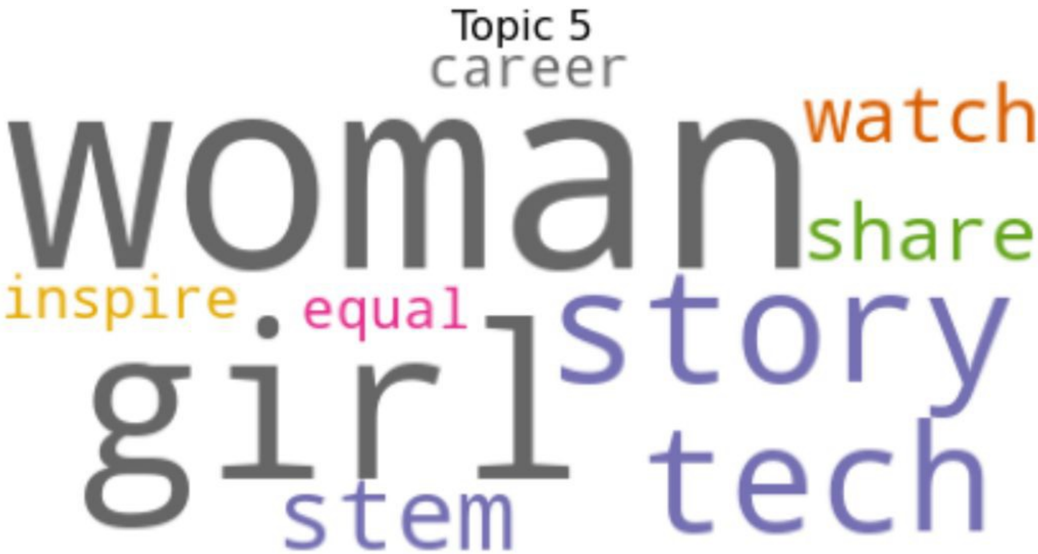
2(c)

Figure 2(c). **Top topics per document and corresponding weight and frequency for Entertainment Industry.** Topics 3 shows distribution with a singular dominant word, and decreasing frequency thereafter.



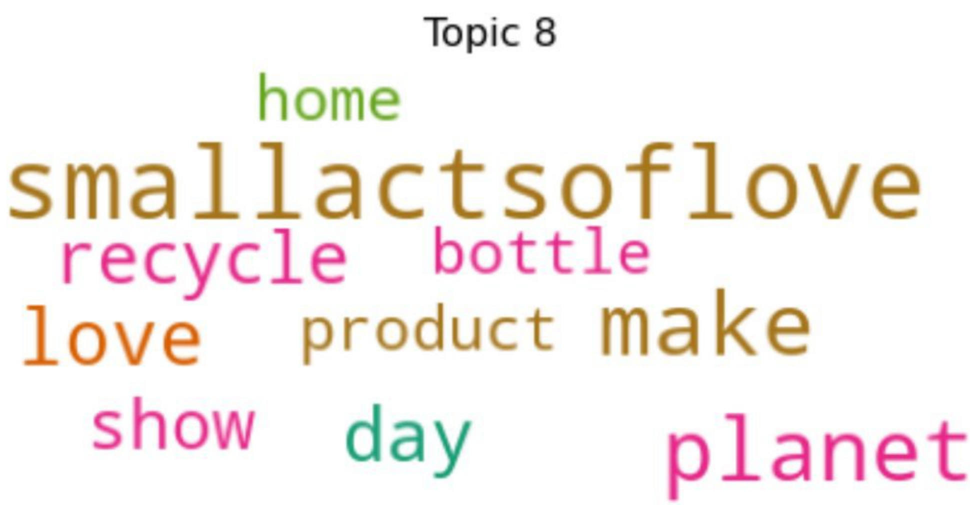
2(d)

Figure 2(d). **Top topics per document and corresponding weight and frequency.** No topic in particular shows word dominance, with frequency and weight varying per each word on all four topic graphs.



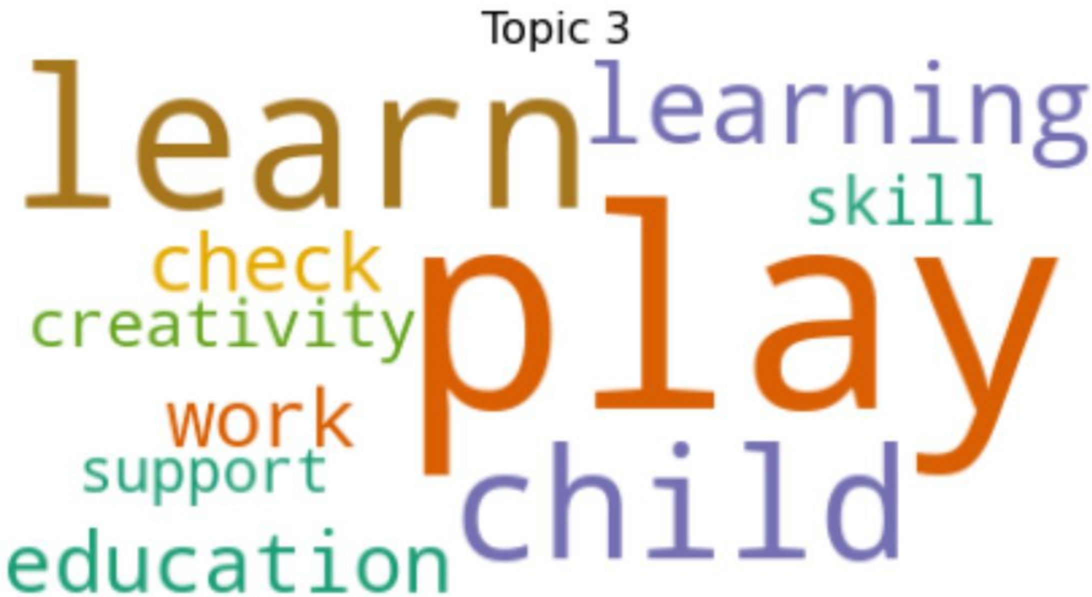
3(a)

Figure 3(a). **Word cloud of most relevant terms for the most salient CSR topic for Tech Industry.** Word clouds represent the word prominence based on size for the top topic per document, displayed for industry groupings only. Figure shows Topic 5, related to social initiatives to support women and girls in stem education and fields. Results for 3((b)-3(d) generated with identical parameters.



3(b)

Figure 3(b). **Word cloud of most relevant terms for the most salient CSR topic for Beauty Industry.** With less visibly dominant terms than figure 3(a), word cloud shows topic 8, which relates to social and environmental initiatives with individual and domestic participation components.



3(c)

Figure 3(c). **Word cloud of most relevant terms for the most salient CSR topic for Entertainment Industry.** Topic 3 is shown with dominant terms representing education and children related initiatives.

3(d)

Figure 3(d). **Word cloud of most relevant terms for most salient CSR topic for Retail Industry.** Topic is represent with large, dominant terms relating to political initiatives including aspects of voter awareness and advocacy.

DISCUSSION

The models and corresponding visualizations of company tweet contents reflected some expected emphases based on industry and account type, while some results showed intense skew and unexpected focus within posted content. Overall, the results show that most companies communicate heavily regarding a few initiatives, most of which fall under environmental and social categories, while almost none fall under economic sustainability. Qualitative analysis showed that periphery to the industry itself was important in which of these categories were most reported on, and that specified CSR accounts tend to exhibit less number of topics, but stronger topics overall. General accounts showed many topics, but most of which were concerned with brand promotion and consumer engagement; however, the beauty industry model specifically showed that relative to the data size, there was a significant impression of CSR related contents across the topic distribution. Other general accounts had little to no topics pertaining to CSR related contents, despite having similar reputations of effective and reputable CSR activity.

Industry Grouping Analysis

Upon observing the pyLDAvis visualizations per industry, several characteristics stand out among the topic distributions. Overall, tech and beauty company communications seem to diversify in topic range more so than the retail and entertainment industries(Figure 1(a)-1(b)). Retail industry company selection may speak for this specific effect, however, as in sampling, green or eco brands were selected on the CSR ranking lists while other industry companies, while not specifically incorporating CSR into their brand purpose or appeal, were still sampled due to their connections to charity, philanthropy, and other CSR initiatives. For many retail brands, internal processes such as manufacturing and production must inherently be reshaped around sustainability to earn brand awareness as a sustainable or ecologically friendly brand, as opposed to the implementation of CSR through external affairs using financial capital. Further, both entertainment and retail topics proved similar across a majority of the document such that the topic chart visualization ended up being dominated by the major, nearly 100% prominence topic(Figure 1(c)-1(d)). Across the board, all industry accounts focused on branding and other content communication before CSR topics; therefore, topics 0-2 for many visualizations did not pertain to sustainability initiatives (Figure 1(a)-1(b)). However, among the displayed CSR topics, the dominant dimensions were environmental for entertainment companies, and social for beauty, tech, and retail companies. While the prominence of environmental topics in the entertainment sector can mainly be attributed to the dominance of DisneyConserve data, for the remaining three industries, it can be deduced that the most common communication of CSR effort involves social good initiatives, particularly those in which consumers can participate or engage with due to product-based sustainability initiatives.

Account Type Analysis

Two of four industry groupings showed account type variation, and thus models were created based on this specification. A given general accounts model was then compared to the designated CSR accounts model within one industry, in order to account for differences that may occur simply due to extremely divergent contents across sectors. Several differences were observed between the account type models, suggesting that within the selected industries, CSR account specialization had an effect on the content topic distribution. Visualizations were not successful and omitted for these groupings, although the LDA models were successful in the script, showing such differences. The specified accounts for entertainment, which included DisneyCSR and DisneyConverves, contained a few salient topics revolving around ecological restoration and conservation, almost exclusively pertaining to environmental sustainability movements. The other two company account, WarnerBros and LegoFoundation, showed models that scattered across brand and event promotion contents, with much smaller and less salient topics pertaining to some childrens' and school-related social initiatives. The two groups within the Technology sector showed little difference; whether or not the company had specific accounts, topics in both social and environmental sustainability were observed. The benefits of specialization therefore seem to depend on the existing topic range within CSR communications; while DisneyCSR severely lacked in data, the data from the newer DisneyConserves and corresponding topic clarity imply that specialization was the favored action in terms of establishing effective CSR communication for this particular company.

Conclusions for Company Social Media Usage for CSR Communication

The findings imply that the method by which certain companies communicate CSR information depend on the industry as well as the dedicated purpose of the account. While these results were to be expected, the nuances in the data highlight the strengths and weaknesses of each selected range of coverage and type of account. Firstly, not all general accounts seem to significantly exempt CSR content, such as within the beauty industry models(FIGURE). Even with a large spread of brand-promotion related topics, content highlighting the commitment to ecological causes and sustainable production was easily identified as one of the top prominent topics. This could possibly be attributed to the younger audiences of prominent beauty companies and the corresponding target content including an activism or wellness lense. Second, it is clear that CSR specified accounts perform better at communicating contents with salient topics, the effect of which may be a more singular brand reputation focusing on activating core purposes. As shown through the entertainment and DisneyConserves data, specificity in account dedication even within CSR tends to create more salient topics, and a strong emphasis on one of the three dimensions -- environmental, economic, or social sustainability(FIGURE). Across all models, economic sustainability was least communicated, whether or not the accounts were generic or CSR dedicated; it seems that there is possibility for growth in this area, although successful consumer interactions and expectations of relatable and commonly understandable social and environmental initiatives may keep this trend constant as online CSR communication progresses forward.

Limitations & Future Directions

The limitations of this study mostly revolve around scale; due to time constraints as well as the manual power involved in cleaning and processing data, even while consulting online scripts, I limited my dataset to a small, focused group of industries and companies. Further, many companies taken into consideration had deactivated or simply removed many previously utilized channels, resulting in a much smaller pool of eligible samples. Other limitations include the adaptation of an LDA model to small texts such as Tweets in general; while many other researchers have utilized this method, due to the nature of tweets severely truncating or limiting otherwise fully communicated ideas in other types of texts, the LDA model is still lacking in alterations that would be better suited to analyze short-form text. Further, in portraying the data, certain analyses failed to create cohesive visualizations due to the lack of data and truncated number of topics. For the industry analyses, topic prominence in document was overshadowed for two industries with the largest topic -- this would require fixing parameters in the code to adjust for smaller percentage values to be included in the chart.

Future research suggestions include a team collaboration to explore the research questions with a larger scale of detail, as well as seeking sample data from companies that may not have made certain CSR rankings in the past few years due their size or net value. Such companies are still likely to active social media platforms that address a smaller, yet more focused and personalized audience, and further analysis could measure topic data against total engagements in order to draw connections between content types and consumer responsiveness. Developments of CCLDA, or Cross-Collection Topic Models, are largely incomplete and have not yet been applied to social media data; the completion of this model may help develop more tools to have quantitative comparisons between two different models, to supplement more number-based analysis to qualitative observations drawn from basic LDA models.

Broader Implications:

CSR communication over social media is the modern wave of reporting sustainable efforts on behalf of firms; observing trends of usage in the past two decades, companies will either join this medium or become more efficient with it. Results of topic modelling across several key industries show that specialization is key in higher coherence within specific models, while topics tend to get lost among the vast amounts of regular brand promotion on the general accounts of highly sustainable companies. As companies move towards improving their presence on social media, CSR initiatives must be effectively communicated to their audiences. Across all industries, there is still room to not only increase the quantity of CSR tweets but to explore the range of their firm's environmental, social, and especially economic efforts. While industry characteristics are likely to rule the topics which are reported, companies within industries like technology exhibit the capability to communicate several key commitments that span all three dimensions within their social media platforms, as could be done by other CSR pursuing companies.

ACKNOWLEDGEMENTS

I would like to thank my family for consistently being there for me and my instructors for supporting me through this process.

REFERENCES

Bicalho, P., M. Pita, G. Pedrosa, A. Lacerda, and G. L. Pappa. 2017. A general framework to expand short text for topic modeling. Information Sciences 393:66–81.

Bowen, H. R. 1953. Social responsibilities of the businessman. University of Iowa Press, Iowa City.

Chae, B., and E. Park. 2018. Corporate Social Responsibility (CSR): A Survey of Topics and Trends Using Twitter Data and Topic Modeling. Sustainability 10:2231.

Cortado, F.-J., and R. Chalmeta. 2016. Use of social networks as a CSR communication tool.

Cogent Business & Management 3:1187783.

Elkington, J. 1999. Cannibals with forks: the triple bottom line of 21st century business. Choice reviews online.

Etter, M., E. Colleoni, L. Illia, K. Meggiorin, and A. D'Eugenio. 2016. Measuring Organizational Legitimacy in Social Media: Assessing Citizens' Judgments With Sentiment Analysis. Business & Society 57:60–97.

Dyllick T, Hockerts K (2002) Beyond the business case for corporate sustainability. Business Strategy and the Environment 11: 130-141.

Freeman, E., and A. Moutchnik. 2013. Stakeholder management and CSR: questions and answers. uwf UmweltWirtschaftsForum 21:5–9.

Flora Hung-Baesecke, C.-J., flora. hung@gmail. com, Y.-R. R. Chen, and B. Boyd. 2016.

Corporate social responsibility, media source preference, trust, and public engagement:The informed public's perspective. Public Relations Review 42:591–599.

Frederick, W. C. 1987. Empirical studies of business ethics and values. JAI PressGreenwich. Fuchs, C. 2009. Information and Communication Technologies and Society. European Journal of Communication 24:69–87.

Global Reporting Initiative. (n.d.). https://www.globalreporting.org/. ICS (n.d.). . http://www.computational-sustainability.org/.

Jhawar, N., and S. Gupta. 2017. Understanding CSR- Its History and the Recent Developments. IOSR Journal of Business and Management 19:105–109.

Koller, M., A. Floh, and A. Zauner. 2011. Further insights into perceived value and consumer loyalty: A "Green" perspective. Psychology and Marketing 28:1154–1176.

Li, S. 2018, June 1. Topic Modeling and Latent Dirichlet Allocation (LDA) in Python Towards Data Science:

https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24.

Mcguire, J. W. 1970. The Future Social Role of the Business Organization. Review of Social Economy 28:9–14.

Lee, K., W.-Y. Oh, and N. Kim. 2013. Social Media for Socially Responsible Firms: Analysis of Fortune 500's Twitter Profiles and their CSR/CSIR Ratings. Journal of Business Ethics 118:791–806.

Mauricio Andrés Latapí Agudelo, Lára Jóhannsdóttir, Brynhildur Davídsdóttir.

A literature review of the history and evolution of corporate social responsibility. International Journal of Corporate Social Responsibility. 2019;(1):1.

doi:10.1186/s40991-018-0039-yxa

Mihalache, S. Ş. 2013. Aspects regarding corporate social responsibility definition and dimensions.

Proceedings of the International Conference Marketing - from Information to Decision 6:130–144.

Natural Language Processing APIs - Twinword API. (n.d.). https://www.twinword.com/api/lemmatizer.php

Grau, S., and Julie Pirsch & Shruti Gupta. 1970, January 1. A Framework for Understanding Corporate Social Responsibility Programs as a Continuum: An Exploratory Study, by JuliePirsch; Shruti Gupta; Stacy Grau. Springer. https://ideas.repec.org/a/kap/jbuset/v70y2007i2p125-140.html.

Porter, M. E., & Kramer, M. R. (2006). Strategy & society: the link between competitive advantage and corporate social responsibility. Harvard Business Review, 84, 78–92

Saxton, M. D. 2018. A Gentle Introduction to Topic Modeling Using Python. Theological Librarianship 11:18–27.

Slaper, T. F., and T. J. Hall. (n.d.). . https://www.ibrc.indiana.edu/ibr/2011/spring/article2.html. Vidal-Leon, C. 2013. Corporate Social Responsibility, Human Rights, and the World Trade

Organization. Journal of International Economic Law 16:893–920.World Business Council for Sustainable Development (WBCSD). (n.d.). .

https://www.wbcsd.org/

Github Repositories: https://github.com/t3abdulg/Twitter-Topic-Modelling
https://github.com/lda-project/lda
https://github.com/hadizh/tweet_topic_modeling?fbclid=IwAR22VoqOu0E2XvRnJdL7XW0GZbW56CaIN8UnNq3kk_ttL-HRBhCNn76A4a4