



Universiteit
Leiden
The Netherlands

Opleiding Informatica & Economie

Exploring Themes of
CEO Social Media Communication

Menno Bruinsma

Supervisors:

Amirhossein Zohrehvand, Guus Ramackers

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

www.liacs.leidenuniv.nl

10-11-2022

Abstract

With the growth of the internet, social media has become an important communication tool for companies. Companies often have their own social media accounts to communicate with the general public and their shareholders. But in addition to these dedicated media accounts, business executives, particularly the CEO, also serve as online spokespersons. CEOs are seen as the face of the company, and their reputation is tied to that of the firm. The goal of this thesis was to examine which topics CEOs of S&P 1500 companies communicate about on Twitter. For this research, 495 CEO Twitter accounts were scraped for tweets, retweets, replies, and citations. The scraped interactions were mapped to vectors with the Python module `Sentencetransformers` which captured both semantics and syntax. We then used the Python module `BERTopic` combined with dimension reduction algorithm `UMAP` and clustering algorithm `HDBSCAN` to group interactions with a similar meaning together. The identified clusters were then manually assigned topics like 'Corporate', 'Politics', and 'Leadership' based on their statistically most significant words. These words gave an accurate description of the content of the cluster. We also researched differences in topics between tweets and retweets using the same methods.

We found that corporate is the most frequent topic that S&P 1500 CEOs communicate about. They mostly use Twitter to share one-way messages about their company/ field. Other popular topics were leadership and politics. We also found that CEOs prefer tweets over retweets when communicating about the topic of leadership. Lastly, CEOs increasingly exhibit more social activism on Twitter, for example about politics and climate change.

The results of this research give a better understanding of general CEO Twitter communication. We identified large-scale trends and behaviors, like the increase in CEO social activism. The methods described in this research can be used as a foundation for future work on large-scale Twitter communication studies.

Contents

1	Introduction	1
2	Related Work	2
3	Methods	4
3.1	Data gathering	4
3.1.1	Finding Twitter handles	4
3.1.2	The Twitter API	5
3.1.3	Scraping Twitter accounts	6
3.1.4	Scraping Twitter interactions	7
3.2	Finding topics with BERTopic	8
3.2.1	Sentencetransformers	8
3.2.2	UMAP	9
3.2.3	HDBSCAN	9
3.2.4	Vectorizer and topic assignment	10
3.3	Hyperparameter tuning	10
3.4	CEO tweet and retweet differences	12
4	Results	13
4.1	All interactions	13
4.2	Tweets	15
4.3	Retweets	16
4.4	Testing the increasing social activism trend	17
4.5	Testing tweet vs retweet topics	19
5	Discussion	20
5.1	Findings	20
5.2	Future work	21
5.3	Limitations	21
5.4	Conclusions	21
	References	22
A	Count of Twitter interactions overtime	26
B	Cluster figures	27
C	Clusters over time	31
D	Twitter interaction topics	32

1 Introduction

With the world becoming increasingly digital, corporate communication changed. Due to social media, CEOs can openly share their thoughts on all subjects whenever they please. In our research, we wanted to examine what CEOs shared on the micro-blogging platform Twitter. By exploring the topics of their tweets, we can detect general communication trends and behaviors. Our findings can potentially improve future CEO social media strategies. Our methods also provide a foundation for further research into CEO Twitter communication.

CEO social media communication has been studied before (e.g., Zohrehvand, 2022). Previous studies found that CEOs often adhere to specific social media strategies (Bendisch et al., 2013; Alghawi et al., 2014; Malhotra and Malhotra, 2016; Heavey et al., 2020). CEOs use Twitter to share information about their company and their personal lives but also to interact with other users (Huang and Yeo, 2018). The previously identified strategies differ mostly on these three goals. Porter et al., 2015 found that CEOs mostly tweeted about their companies. We expect to find this result in our research as well.

We wanted to perform a large-scale analysis of CEO social media communication. We chose Twitter since a large number of CEOs have an account and the Twitter API is well-documented. In our research, we applied the newest topic modeling methods; therefore, our results could potentially be more robust compared to the few studies that also performed tweet scraping combined with topic modelling in the past. To examine CEO communication, we used a data set composed of S&P 1500 CEOs. Because this data set contains CEOs from small, medium, and large market cap sizes, it gives an accurate insight into the communication of the average CEO. In the data set were 495 CEOs who have a Twitter account.

The goal of this paper was to examine the topics of CEO communication on Twitter. We also wanted to determine whether the two main types of interaction, tweeting and retweeting, differ in topics. In order to research this we stated the following two questions:

1. What topics do CEOs communicate about on Twitter?
2. What are the differences between CEO tweet and CEO retweet topics?

To answer the first question, we gathered all the Twitter interactions ever made by the CEOs and applied an unsupervised clustering algorithm to these interactions. The interactions we looked at are tweeting, retweeting, replying, and citing. We manually assigned topics to the clusters based on their statistically most significant words. When clusters are mentioned in the paper we refer to clusters that are found by the clustering algorithm itself. When topics are mentioned, we refer to the topics that we ourselves assigned to the clusters. The second question also made use of the clustering algorithm, but we now ran the algorithm separately on tweets and retweets. We looked at both statistical differences and empirical differences. Because a large number of Twitter accounts were researched, we did not specifically look at what individual CEOs share or research what their individual social media strategies are.

2 Related Work

Social media users with a large number of followers are also called influencers. More precisely, Cambridge University Press, n.d. defined influencers in business English as: “A person or group that has the ability to influence the behavior or opinions of others”. Influencers can range from movie stars to beauty vloggers, but are most often seen as regular people who amassed a large social following due to their online content (Lou and Yuan, 2019). Nowadays, firms often utilize these influencers for their marketing campaigns in order to promote products or services (Childers et al., 2019). Since CEOs generally represent their company in the media, be it traditional mediums like newspapers and TV or platforms like Twitter and Facebook, we argue that CEOs can also be seen as a type of influencer. While many do not have the same following size as the average influencer, especially CEOs from smaller companies, their active social media presence does have effects on their own firm (Shandwick, 2012). The best and most famous example of a CEO who acts as an influencer is Tesla CEO Elon Musk. With a following of 102 million Twitter users at the time of writing, his range of influence is enormous. His past tweets influenced both Tesla stock prices and the cryptocurrency market (D. Kim et al., 2021; Ante, 2021).

However, many CEOs are not active on social media. Most CEOs do not have a Twitter account and the ones that do are generally not very active (Capriotti and Ruesja, 2018). Reasons for this are amongst others time and risk (Fischer and Reuber, 2011; Shandwick, 2012; Malhotra and Malhotra, 2016). The fear of posting something that is viewed as bad by social media users is enough to deter CEOs from using platforms like Twitter. Accounting and auditing firms often advise companies on risk management regarding social media. They argue that abstaining from social media is the worst risk since you can’t quickly respond to potential issues (Elliott et al., 2018). A good response on social media to problems or complaints can increase the social perception of the company (Xia, 2013). A rapid public apology by the CEO following a company crisis or problem greatly reduces negative sentiment towards the company (H. Kim et al., 2015). This strongly suggests that all CEOs should be active on social media. Although the risks of turning consumers against you due to a bad communication strategy definitely exist, CEOs should not abstain from social media. CEOs do not have to necessarily engage in discussions with the public on the platform or communicate their own opinion. Social media can still be used as a one-way communication platform for company-related messages. This way, CEOs do have an online presence and can still quickly react if necessary, without bearing much risk. Due to the potential risks and the significant influence of CEO image on firm reputation we make the following hypothesis.

Hypothesis 1: *CEOs mainly use Twitter to communicate about their own companies.*

Tweeting about your own company’s services, products, or employees is less risky than tweeting about personal opinions and criticizing others. Twitter is in this case used for one-way communication with its users.

While less risky, tweeting about your own company is less interesting for social media users (Emerald, 2018). Tweeting about more personal and insightful topics will likely attract more followers since users are interested in the leadership capabilities and success of CEOs (Men and Tsai, 2016). CEOs could use personal branding to differentiate themselves from each other. We define personal branding as purposefully creating a public perception of an individual with certain character traits, opinions, capabilities, and achievements. Osnat, 2019 stated that: “CEO brands have become the most significant components of the company’s corporate reputation and brand”. CEOs can build their own personal brand by sharing information and opinions about specific topics or engaging with others users. For example, to build customer loyalty and awareness, executives should engage with social media users and be more than just corporate figures (Karaduman, 2013). Over the years multiple studies have identified different CEO social media strategies (Bendisch et al., 2013; Alghawi et al., 2014; Malhotra and Malhotra, 2016; Heavey et al., 2020). All these strategies made use of personal branding in some way.

An effective way to build a strong personal brand is by conveying social activism. CEOs increasingly discuss political and social issues like immigration, climate change, and equal rights on social media (Larcker et al., 2018). CEOs have never before shown this much activity and interest in politics online (Wolfe, 2019). Some CEOs even openly shared their political views and even endorsed their preferred candidate for the 2020 US election (Bondi et al., 2022). This social activism mostly comes from personal values but also from expectations of support from employees and customers (Hambrick and Wowak, 2021). A portion of CEOs also see it as a way to gain a competitive advantage (Bielak et al., 2007). When a CEO takes a stand on a controversial subject, he/she can rally support for the company from users who share the same view. However, this activism risks losing support from consumers who disagree with the stance of the CEO (Chatterji and Toffel, 2019). We suspect that CEOs mostly side with the opinion of their own target segments to avoid alienating large groups of consumers. A stance we would expect CEOs to currently take is that climate change is real and that we need to take action X or Y to prevent the problem from getting worse. However, not all CEOs approve of sharing political views and some even take an apolitical stance (Bondi et al., 2022). We expect to recognize this increase in social activism in our research and state the following hypothesis.

Hypothesis 2: *CEO social activism shows an increasing trend on Twitter.*

CEO communication can also be categorized within a platform. In the case of Twitter, tweets are a way of directly communicating with the general public, consumers, and shareholders. Retweeting has a more indirect nature since you re-share a message from someone else. We expect that will CEOs tweet more about less risky topics like leadership and corporate. In turn, we expect that retweets will contain more controversial topics like politics. Thus, we suggest that there is a difference in topics between these two interactions. We state the following hypothesis.

Hypothesis 3: *There is a difference in topics between CEO tweets and retweets.*

3 Methods

This section discusses the methods that we used to collect our results. First, the processes of gathering Twitter accounts and Twitter interactions are covered. Second, we discuss the BERTopic clustering process and the topic assignment of identified clusters. Third, we talk about setting hyperparameters for our different algorithms. Lastly, the statistical test for finding the difference in topics between tweets and retweets is discussed¹.

3.1 Data gathering

The gathering of Twitter interactions was conducted in three different steps. First, we searched for the Twitter handles of the CEOs. After this, we scraped the Twitter user IDs of the identified accounts using the Twitter API. With these user IDs we scraped all the interactions (tweets, retweets, replies, citations) made by these accounts, again making use of the Twitter API.

3.1.1 Finding Twitter handles

As previously mentioned, we used a data set made up of CEOs from the S&P 1500. In this data set were 5500 current and former CEOs of companies from this stock index. The data set contained information such as name, company, executive ID, join date and leave date. The data set was obtained from ExecuComp, through WRDS². WRDS is a data platform with a lot of data sets about finance and management. WRDS is used by companies and research institutions from all over the world.

In order to find the Twitter accounts of these CEOs, we had to manually search each CEO on the internet³. Sometimes it was unclear if an account actually belonged to a CEO. For example, there are sometimes fake accounts made by angry customers to smear the image of the CEO and the company he/she works for. CEOs also change companies, which makes it more difficult to verify that it is actually the same person. With these harder cases, we had to look at photos and other social media accounts like LinkedIn to verify that it is the same person. To find the CEOs, we used search queries where we included the full name and company, followed by ‘Twitter’.

An example of a search query we used would be:

Jeffrey P. Bezos AMAZON.COM INC Twitter

¹Our code can be found at <https://github.com/MennoBruinsma/Bachelor-Thesis-CEO-Twitter-Communication>.

²This data was received from research supervisor Amirhossein Zohrehvand. This research is part of a larger social media research project, which he leads.

³This was done together with fellow student Dylan Macquine who was also part of the larger project.

Of the 5500 CEOs in our data set, 495 actually had a Twitter account. This number is low but this was an expected result. In our data set, CEOs were predominantly male between the ages of 45 and 65. People in this age range are generally less active on social media (Mellon and Prosser, 2017). Also, many CEOs have no time for or even fear the use of social media (Fischer and Reuber, 2011; Shandwick, 2012). There is a significant difference in activity between these accounts. Some post almost daily and others have not shown any activity in years. However, this is not a problem for our research since we were interested in the topics CEOs communicate about and not in their activity.

3.1.2 The Twitter API

With all Twitter handles identified, we could proceed with the next step: scraping the Twitter user IDs of the accounts. To do this we required access to the Twitter API. The Twitter API has multiple types of developer tiers. The standard free developer account is allowed to scrape 100,000 interactions per month. Since each tweet, retweet, reply, or citation uses one call this is not enough to scrape all accounts and their corresponding interactions. That is why we used an academic developer account which is allowed ten million interactions per month.

During this research, the newest version of the Twitter API was used: Twitter APIv2. From now on if we mention the Twitter API we actually talk about Twitter APIv2. The Twitter API can also be used for more purposes than just scraping. For example, the Twitter API enables auto-following, auto-tweeting, auto-deleting, and much more (Twitter, 2022).

Each developer account comes with API keys. These keys give you access to Twitter when you make an HTTP/HTTPS request. Twitter enables developers with an academic account to scrape everything except personal messages. For our research, we were only interested in tweets, retweets, replies, and citations. We used Python to scrape Twitter and then particularly the modules `searchtweets v2`, `pandas`, and `request`. The version of `searchtweets v2` that was used is solely made for academic Twitter accounts and can not be used with by any other account type (Zohrehvand, 2021). The regular `searchtweets v2` version can be used by premium and enterprise developer accounts as well. `Pandas` is one of the most popular Python modules for handling data. It allows for easy and quick manipulations of large quantities of data. The `request` module enables Python to make HTTP requests to web applications.

3.1.3 Scraping Twitter accounts

With the API keys and searchtweets v2 we were able to make the requests to scrape the Twitter IDs from the CEOs. To make a request you need to give an endpoint, parameters, and authorization keys. An endpoint defines to which web application you want to make a request. An endpoint makes use of HTTP or HTTPS calls to access information on a web application. An example of an endpoint we used to scrape the Twitter accounts would be:

```
https://api.Twitter.com/2/users/by?usernames=jeremys,kaufers,mikegianoni
```

The parameters define what you expect from Twitter to return in the response. This can range from the user ID to the URL of the profile picture. The parameters we used to scrape the Twitter accounts are listed in Table 1.

With the endpoint, keys, and parameters we could make the requests to the Twitter web application. Twitter returned the response back in JSON format. We merged all the converted JSON responses into a CSV file. This way, we only needed to scrape the accounts once.

Parameter	Description
id	The unique identifier of the account.
username	The username of the account.
name	The name defined by the user, is not always equal to the actual name.
created_at	Creation timestamp of the account.
verified	Indicates if the account is verified.
description	The bio of the account.
profile_image_url	The URL to the profile picture of the account.
pinned_tweet_id	ID of the tweet that is pinned by the account.
public_metrics	Information about the count of followers, following and tweets
url	The URL on the profile specified by the account
context_annotations	Contains context annotations for the tweet
location	The location of the account specified by the user.
withheld	Contains information if the account is being withheld in some countries
protected	Indicates if the account is closed for non followers

Table 1: Used parameters to request data about CEO Twitter accounts

3.1.4 Scraping Twitter interactions

With the user IDs stored in a CSV, we were now able to scrape the interactions of these accounts. It works in a similar way as scraping the Twitter user IDs. In order to make the requests, we again needed an endpoint, keys, and parameters. The endpoint for scraping interactions of one account is:

”https://api.Twitter.com/2/users/id/tweets”.format(id = user_id)

The main difference with the endpoint for the accounts is that we call one account here. This is because users can have a large number of interactions and the API response has a limit in size. We looped through each row in the CSV and make a request for every CEO. For each interaction, we requested the parameters as seen in Table 2. It is possible to pass more parameters but these require permission from the account that tweeted and were also not required for this research. Since we made a large number of requests to Twitter, we waited two seconds after each request before we made a new one. This is common practice with web scraping and prevents overwhelming web applications with API requests. We again merged all responses into a CSV for future use, so we would not need to scrape again. The tweets were scraped on 03/04/2022, so all tweets, retweets, citations, and replies made on and before this day by the CEOs were included. The counts of the scraped interactions are shown in Table 3. As seen in Table 3, CEOs prefer to use tweets and retweets in their communication. The evolution of the interactions over time can be found in Appendix A Figure 9.

Parameter	Description
ID	The unique identifier of an interaction
author_id	The unique identifier of the account that made the interaction
created_at	Creation timestamp of the interaction
text	The UTF-8 text in the interaction
referenced_tweets	Shows parent interactions, this means that this interaction is a reply, retweet or citation
public_metrics	The public engagement of the interaction at the time of the request
lang	The language of the interaction
entities	Extra entities about the interaction like hashtags and URL’s
conversation_id	The ID of the original parent interaction
in_reply_to_user_id	The Twitter ID of the user the tweet was replying to
context_annotations	Contains context annotations for the interaction
attachments	Contains the type of attachment like videos or images
withheld	Contains information if the interaction is being withheld in some countries
source	The name of the application used to post the interaction

Table 2: Used parameters to request data about Twitter interactions

	Tweets	Retweets	Replies	Citations	Total
Count	191388	106712	48541	35202	381843

Table 3: Count of the different kinds of interactions scraped on 03/04/2022

3.2 Finding topics with BERTopic

To find the topics of the CEO Twitter interactions, we used an unsupervised clustering algorithm. Since cluster algorithms work with vectors, we needed to transform the textual interactions into vectors. With these vectors, we could perform the clustering and topic assignment. The steps in this process are also shown in Figure 1. This entire process is simplified by using BERTopic. BERTopic is a Python module that makes use of the most recent clustering methods and algorithms to find topics in large sets of documents (Grootendorst, 2022). It allows multiple different embedding, dimensional reduction, and clustering algorithms of your own choice. BERTopic returns clusters, cluster descriptions, and cluster visualizations. However, BERTopic does not assign topics to the found clusters. The topics were manually identified by interpreting the meaning of the statistically most significant words of a cluster (Marchetti and Puranam, 2020). In this section, the different intermediate steps that BERTopic takes are explained.

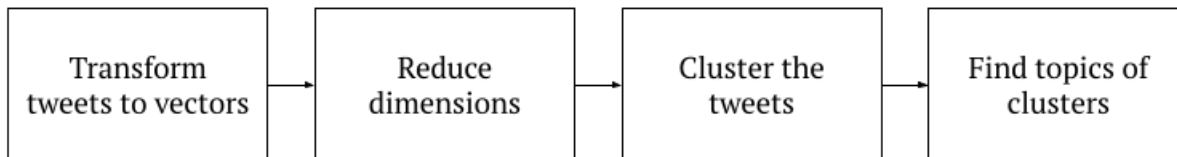


Figure 1: Steps of finding topics in scraped tweets

3.2.1 Sentencetransformers

The first step in the process was to transform the scraped interactions into vectors. To achieve this we used the sentence-transformers module and its SentenceTransformer function (Reimers and Gurevych, 2019). It gives the option to choose from multiple pre-trained sentence embedding models. A sentence embedding model maps sentences to a vector space and captures the semantics and syntax of the mapped sentences. Sentences that have the same meaning should be mapped close to each other in the vector space by the model. We used *all-MiniLM-L6-v2* as our sentence embedding model because it runs fast and still delivers a good embedding performance. There are other pre-trained models that could potentially perform better, but these would take a lot longer to run. *All-MiniLM-L6-v2* is very well suited to embed sentences and short paragraphs into vectors and is trained on a data set of over one billion sentences. The model has a limit of 256 words, all words after this limit will be truncated. Since tweets have a limit of 280 characters, this did not pose a problem⁴. The model maps sentences to a 384-dimensional vector space. The process of mapping the sentences to vectors takes a long time. This is why we pre-computed the embeddings and saved them to Numpy files for later usage. This reduced the time of testing the hyperparameters significantly.

⁴The model is only trained on English sentences. While most interactions in our data set are in English, a small number (1.3%) are in a different language. The model is thus not able to properly map these interactions to the correct place in the vector space.

3.2.2 UMAP

Clustering algorithms often struggle with high-dimensional data. Although our data was not high-dimensional per definition (a data set with more dimensions than rows), it still had 384 dimensions. The clustering algorithm we chose works better when there are fewer dimensions. To reduce the number of dimensions, the Python module UMAP was used. UMAP stands for Uniform Manifold Approximation and Projection for Dimension Reduction. UMAP first constructs a high-dimensional graph representation of the vector space. When this is finished, UMAP will make a similar low-dimensional graph that has the same structure (McInnes et al., 2018). We chose UMAP because it captures the local and global structure of the data. It also allows for choosing the number of dimensions you want to reduce it to. With the use of UMAP, the dimensions were reduced from 384 to five dimensions. We chose five because with this number our clustering algorithm worked both fast and accurately. The fewer dimensions the vector space is reduced to, the more information about the original vector space gets lost. By reducing the amount of dimensions to a larger number, for example ten, the clustering algorithm will have a significantly longer run time. During our testing, we found little to no differences except the run time when we used more than five dimensions.

3.2.3 HDBSCAN

For the clustering itself, we used HDBSCAN (Campello et al., 2013). HDBSCAN is a hierarchical clustering algorithm that makes use of the spatial difference between data points in a vector space. First, it tries to find dense locations in the space. These high dense areas can be seen as mountains on the land, and we wanted to identify these mountains. This is done by first measuring the mutual reachability distance. This is the maximum of the core distance of point x , the core distance of point y and the distance between x and y . The core distance is the distance of a point to its K -th nearest neighbor. The data points in denser regions have a lower mutual reachability distance than points in sparse regions. With these distances a minimum spanning tree is constructed with Prim's algorithm. Each edge in the tree has the weight of its the mutual reachability distance. Next, a cluster hierarchy based on the minimum spanning tree is created. Then, the tree gets reduced by using the user defined parameter Minimum Cluster Size. The algorithm walks through the hierarchy from the root and determines at every split if the children have less data points than the minimum cluster size. If one of the two has less then this is seen as points falling out a cluster. The algorithm notes which points fell out at which distance and uses the information later to determine the final clusters. If both children are large enough a regular split happens and both children are seen as valid clusters. HDBSCAN then chooses the best clusters from this finalized condensed tree.

HDBSCAN is based on the spatial clustering algorithm DBSCAN (Ester et al., 1996). DBSCAN was the first clustering algorithm that did not force all data points to a cluster. This means that the accuracy of the clusters is higher. HDBSCAN works mostly in the same way as DBSCAN but does not need a hyperparameter for the minimum distance between points like DBSCAN does. This parameter is challenging to tune and HDBSCAN takes this process away. OPTICS is another algorithm that is based on DBSCAN (Ankerst et al., 1999). However, OPTICS still requires a parameter that defines when a region is dense. This parameter is also difficult to tune. HDBSCAN only requires the minimum cluster size as a parameter to work properly, which makes it easier to tune/use than the other algorithms mentioned before. HDBSCAN focuses on making clusters as accurate as

possible and leaving data that it is unsure about out. This means that actual clusters might be larger.

HDBSCAN does not require a parameter which specifies the amount of clusters that need to be returned. It determines the number of clusters by itself based on the minimal cluster size parameter and the provided data. This means that the number of clusters HDBSCAN returns cannot be controlled precisely. The only way to raise or lower the number of clusters that will be returned is to use the minimum cluster size parameter. Data points that did not get assigned to a cluster by the algorithm are treated as noise and will be left out.

3.2.4 Vectorizer and topic assignment

BERTopic makes use of a modified version of the TfidfTransformer from Sklearn called c-TF-IDF to find the most important words for the identified clusters (Pedregosa et al., 2011). This is done by first merging all the documents (in our case Twitter interactions) that are in a cluster into one large document. To then find the importance of a word in a cluster, BERTopic (Grootendorst, 2022) uses the formula:

$$W_{x;c} = tf_{x;c} \log(1 + A/f_x)$$

$tf_{x;c}$ = the frequency of word x in cluster c

f_x = the frequency of word x in all clusters

A = the average number of words per cluster

For each cluster, the ten most important words are computed using this formula. Like regular TF-IDF, the size of the weight corresponds with the importance of a word in a document, in this case a cluster. These ten words often give an understandable description of a cluster. This made it possible to manually determine the topics of the found clusters. For instance, a cluster on US politics will have words such as ‘Trump’, ‘Biden’, ‘US’ and ‘election’. We could then give the cluster the topic ‘Politics’. Similarly, a cluster talking about company products, services, values, or jobs will be assigned the topic ‘Corporate’.

3.3 Hyperparameter tuning

BERTopic allows for multiple hyperparameters to be passed when calling the module. The most important parameter is Minimum Topic Size. This determines the minimum size of the generated clusters. Setting the value too high can result in no clusters, and setting it too low results in way too many. Since we worked with three data sets of different sizes (All interactions, tweets, and retweets) we used different values of this parameter. The final values were found by manually trying different values and looking at the sizes and important words of the found clusters. When passing a larger value for Minimum Topic Size, larger clusters are formed. This is because some clusters now include Twitter interactions that were originally in different smaller clusters. When passing a small value, lots of clusters get formed and more data points are assigned to a cluster. However, we wanted to look at large general topics, which means we wanted large clusters. See also Table 4 for the final values.

Another important hyperparameter is Number Of Topics. BERTopic will return exactly the specified number of clusters you requested with this parameter. BERTopic built this as an extra feature for HDBSCAN, because like stated before, HDBSCAN does not allow you to do this. The danger is that clusters that are very different in meaning will be merged together to create the specified number of clusters. We do not want this and thus always set this value to None.

N Grammatical Range and Top N Words both have an effect on how the vectorizer works. N Grammatical Range sets the length of words that will be looked at. For example, with [1,2] words like New York are included while with [1,1] they are not. Top N Words is the number of important words that are found for each cluster. Most often only the first ten words are actually relevant for a cluster. Generating more would only increase run time while little value is added.

The hyperparameters described above were tuned manually. We took all the default parameters of BERTopic and tuned them to see if our clusters would change in size/important words combined with an increased/reduced run time. The values in Table 4 yielded the most interpretable clusters with good cluster sizes. UMAP and HDBSCAN hyperparameters could also be changed but we used all the default values passed by BERTopic. When testing, the default parameters performed the best on our data. Hyperparameter tuning on cluster analysis with UMAP is challenging since the vectors get mapped differently by UMAP on every run. The results are similar but never exactly the same. It is possible to make the dimension reduction consistent with the Random State parameter of UMAP. However, this decreased performance and led to smaller clusters when we were testing. The unsupervised nature of clustering also makes tuning difficult since it is not known how large the clusters should be. There is no real right or wrong answer with unsupervised clustering.

Hyperparameter	All interactions	Tweets	Retweets
Minimum Cluster Size	600	400	300
Number Of Topics	None	None	None
N Grammatical Range	[1,1]	[1,1]	[1,1]
Top N Words	10	10	10

Table 4: Hyperparameters we used with BERTopic

3.4 CEO tweet and retweet differences

To research whether there is an actual difference in topics between tweets and retweets, we wanted to perform a statistical test to compare the means of the two samples. BERTopic gives the option to return a tuple containing the document ID and the cluster number it was assigned during the clustering. We could rejoin the document IDs on the tweet and retweet CSVs to match cluster numbers to individual CEO tweets and retweets. The manual topic assignments made before, which were based on the ten most significant words of a cluster, were also used here. Each general topic acts as a parameter for both samples. The Twitter accounts that never made a tweet or a retweet were left out since these did not contribute anything meaningful. The number of CEOs who tweeted or retweeted was 470.

With this method, we had two samples with multiple parameters (the topics). We decided to use a paired T-test on each parameter. A paired T-test checks for equality of means for two dependent samples. For each T-test we made the following assumptions:

1. The data is a random sample of the population.
2. The data follows a normal distribution.
3. The data is dependent.

If there were no difference in topics between the two samples, each parameter would have the same mean as the parameter in the other sample. For each test we made the following hypotheses:

1. Null hypothesis (H_0): $t = r$
2. Alternative hypothesis (H_A): $t \neq r$

We test with a p-value of 0.05. If the p-value from the test is less than 0.05 the null hypothesis will be rejected and the alternative hypothesis will be accepted. If the p-value is larger than 0.05 the null hypothesis will be accepted. The tests were performed with Python and the scipy module (Virtanen et al., 2020). Scipy is a scientific computation library that provides more utility functions for optimization, statistics, and signal processing. Since we were searching for differences in topics, one rejection would be enough to state that there is a statistically significant difference between the two samples and thus between the topics of the interactions.

4 Results

In this chapter, the results of the clustering method described in Chapter 3 are stated. We looked at the clusters of all Twitter interactions, clusters of tweets, and clusters of retweets. We manually assigned each cluster a general topic based on its most important words. Secondly, we performed a statistical test to research the increasing social activism trend. Lastly, we show the results of our statistical test on the difference in topics between tweets and retweets.

4.1 All interactions

The CEOs in our data set have made 381843 Twitter interactions since 2008. We ran BERTopic on these interactions and clustered them using the methods described in Chapter 3. Of the 381843 interactions, 175635 (46%) were assigned to a cluster by the algorithm. These interactions formed a total of 72 unique clusters in the vector space. This number was determined as the optimal amount by HDBSCAN based on our passed parameters. The interactions that did not get assigned to a cluster are treated as noise by the algorithm and were left out. To visualize the clusters we used a two-dimensional representation of the vector space, as seen in Figure 2. In this two-dimensional environment, several clusters are close to or overlap one another. This often means that they share the same subject, like politics or corporate. For example, the group of clusters in the middle of the bottom right quadrant in Figure 2 have the general topic 1 ‘Corporate’. These clusters have interactions that talk about space travel and satellites. However, sometimes these overlapping clusters had little or nothing in common. This is because a two-dimensional space does not adequately capture the complexity of the original space.



Figure 2: Clusters found in all interactions, closeness generally represents similarity

In Figure 3 the cosine similarities between the different clusters are shown. A similarity score of 1 means that the clusters were identical and a score of 0 entails that they had nothing in common. The clusters found by the algorithm were often still too detailed for our research objectives. The cosine similarity between clusters that are actually similar in their general topic was often low according to the matrix. However, the algorithm was able to produce understandable descriptions of the clusters found by providing the most significant words. This enabled us to manually assign the general topics.

We found that the topics ‘Corporate’, ‘Leadership’, and ‘Politics’ were the most communicated topics. For the sake of simplicity, we put all other identified smaller topics under the general topic ‘Other’. The sizes of these topics are shown in Table 5. A more precise list of topics and their sizes can be found in Appendix D Table 9. Corporate is the largest topic identified, making up 43% of interactions that were assigned a cluster. This finding supports hypothesis 1, which stated that CEOs mostly communicate about their own companies.

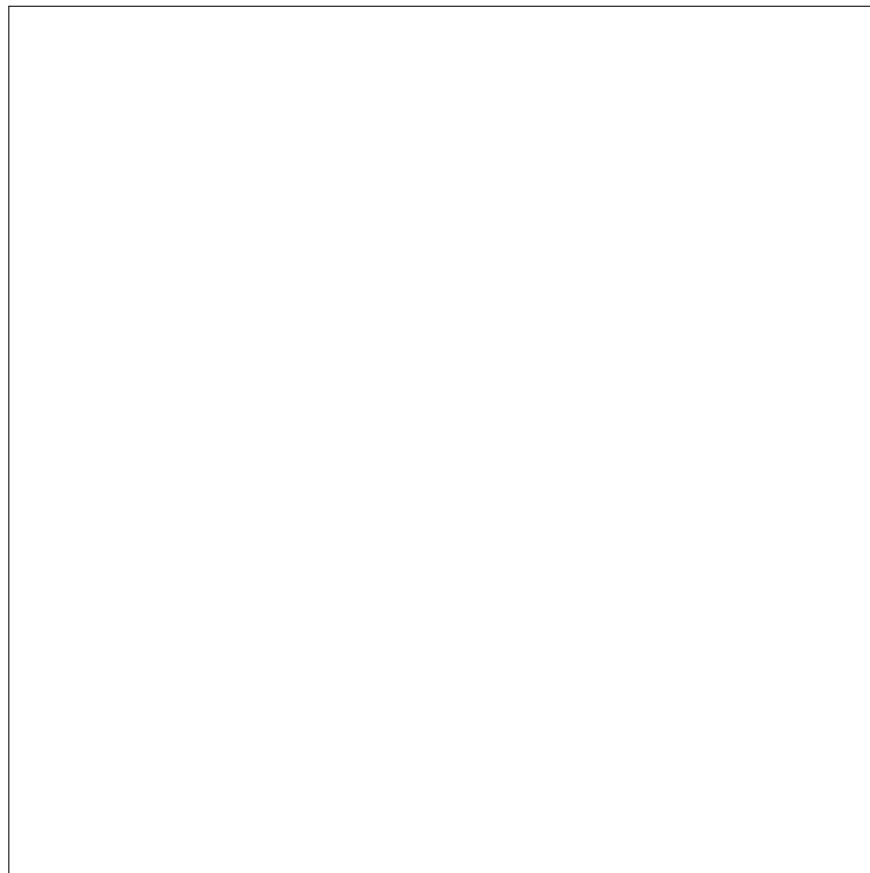


Figure 3: Cosine similarity matrix all interactions

	Corporate	Politics	Leadership	Other	Total
Size topic	75379	19710	15999	64277	175635

Table 5: Sizes of the largest topics for all interactions

4.2 Tweets

The CEOs have made a total of 191388 tweets on Twitter. We applied the exact same methods as with the clustering of all interactions. The intertopic distance map and the similarity matrix for the tweets can be found in Appendix B. Of the 191388 tweets, 87534 (46%) were assigned to a cluster by the algorithm. There were 59 different clusters identified in the vector space. We performed manual topic assignment on the clusters based on the most important words. The topics ‘Corporate’, ‘Politics’ and ‘Leadership’ were again the largest. Smaller topics were again merged under the topic ‘Other’. See also Table 6 and Appendix D Table 10. Since we also scraped the dates of creation for the Twitter interactions (see Table 2), we knew exactly when each tweet was made. With this data, we could plot the evolution of the four topics over time in Figure 4. We removed the 2022 data points for the plot because with only three months of data (January, February, and March) it would skew the graph. The plot suggests that the CEOs are increasingly communicating more about the smaller topics when tweeting.

	Corporate	Politics	Leadership	Other	Total
Size topic	27758	9999	13295	36482	87534

Table 6: Sizes of the largest topics for the tweet interaction

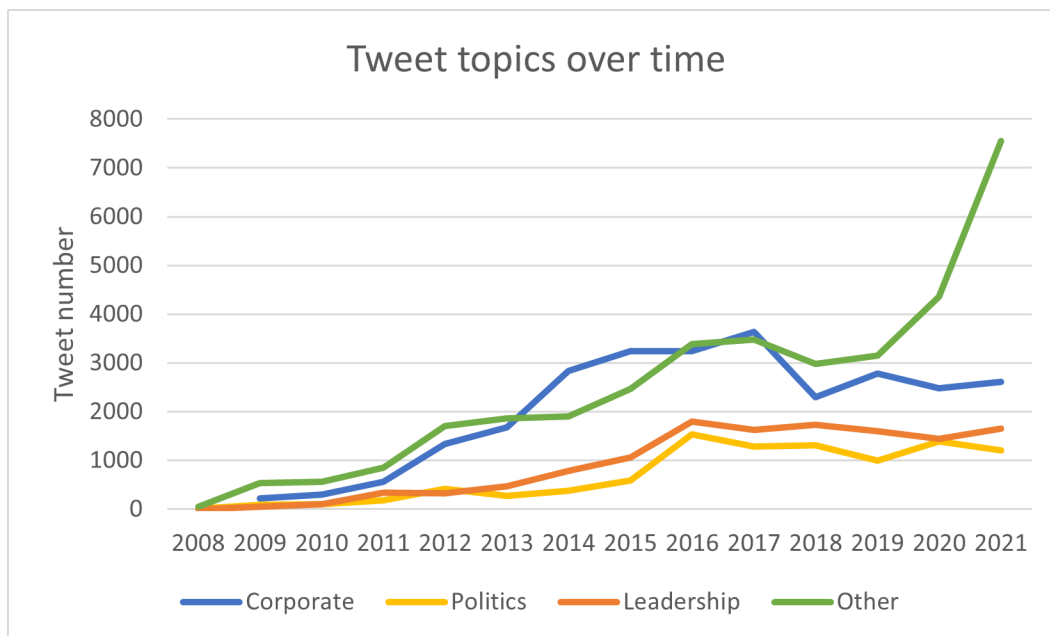


Figure 4: Largest topics of tweets over time

4.3 Retweets

The CEOs have made a total of 106712 retweets from 2008 till 03/04/2022. We applied the exact same methods that were used before. The intertopic distance map and the similarity matrix for the retweets can be found in Appendix A. Of the 106712 retweets, 53288 (49%) were assigned to a cluster by the algorithm. There were 49 different clusters identified by HDBSCAN. Again, ‘Corporate’ was the largest cluster. However, the topic ‘Sports’ was slightly bigger (by eleven retweets) than the topic ‘Leadership’. To keep the different interactions comparable, we consider ‘Leadership’ as one of the three largest topics for retweets, since the difference between the two topics is so minor. See also Table 7 and Appendix D Table 11. We also plotted the evolution of the topics for retweets, as can be seen in Figure 5, again without 2022 data. It is interesting to see that according to our model the topic ‘Corporate’ peaked in 2017 and then started to decline in popularity. Another interesting observation is the peak for the topic ‘Politics’ in 2020, the year of the US presidential election.

	Corporate	Politics	Leadership	Other	Total
Size topic	24015	7680	4368	17225	53288

Table 7: Sizes of the largest topics for the retweet interaction

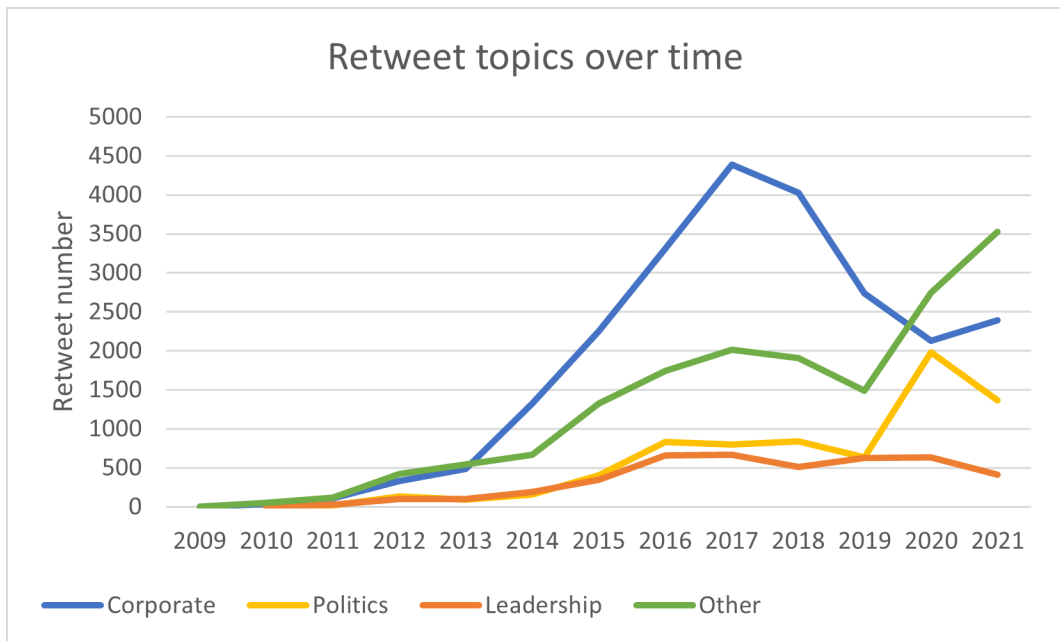


Figure 5: Largest topics of retweets over time

4.4 Testing the increasing social activism trend

We were also interested in how social activism evolved over time and if there is an increasing trend. We first plotted the largest clusters found in all the interactions to see if we could spot any interesting trends, see Figure 6. We left the 2022 data out because we only had three months of data. If we had included 2022, the graph would be skewed and would give a wrong impression of the Twitter activity. This plot allowed us to see certain clusters spike in activity, for example, COVID in 2020. When looking at Figure 6, three of the five largest clusters were about social issues. Both COVID and the US election saw a large growth in activity in 2020, and sustainable energy in 2021. This plot suggests a potential increasing trend of CEO social activism.

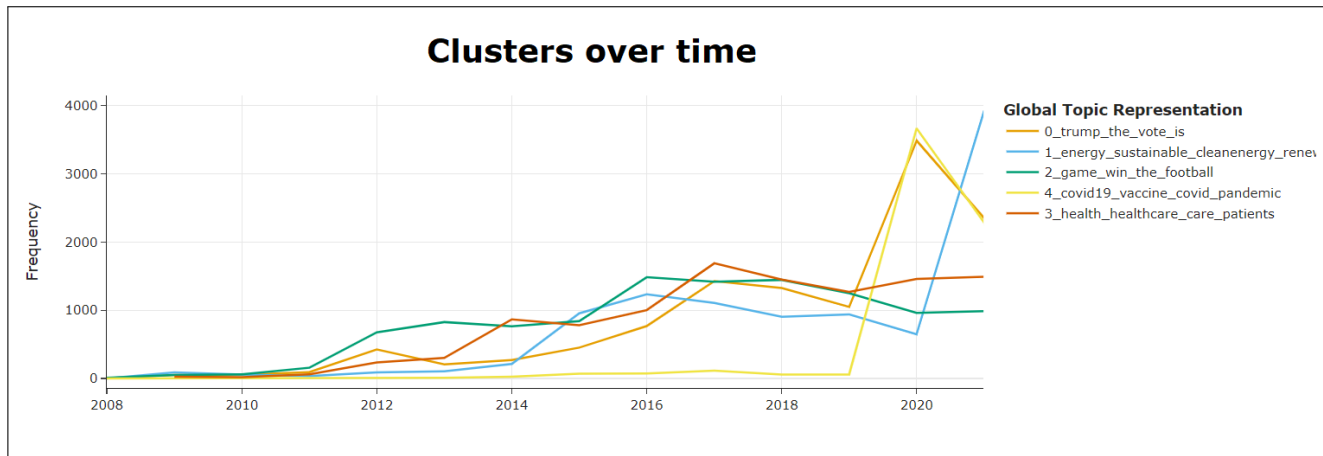


Figure 6: Largest clusters of all interactions over time

To actually test the trend of increasing CEO social activism stated in hypothesis 2, we chose to first perform Kendall's Tau to measure the strength and direction of the trend. Kendall's Tau is a measure of correspondence between two different time series (Ostwal, 2020), in our case the years and the number of tweets over the years. A value of 1 means a strong correspondence, and a value of -1 means a strong disagreement. We performed the test using the SciPy library (Virtanen et al., 2020). We added the three clusters that indicate social activism from Figure 6 (cluster 0, cluster 1, and cluster 4) together and performed the test. The resulting Tau was 0.85, which indicates a strong positive correlation between time and the number of social activism tweets. However, this does not count as statistical evidence. Therefore, we performed the following hypothesis test with a p-value of 0.05:

1. Null hypothesis (H_0) : $T = 0$
2. Alternative hypothesis (H_A) : $T \notin 0$

We got a p-value of less than 0.00000, which means that the null hypothesis was rejected and the alternative hypothesis was accepted instead. This means that the increasing trend does indeed exist.

We also performed linear regression with the same data using scikit-learn (Pedregosa et al., 2011). We used years as a way to predict the amount of social activism interactions. We got a R squared of 0.66 and a slope of 537.8, see also Figure 7. The positive sign of the slope indicated a positive relationship between years and the number of social activism interactions. The R squared means that there is a medium correlation between the two variables. The regression line also had a p-value of 0.0003, which means that years have a statistically significant positive influence on social activism interactions.

However, the amount of interactions over time has grown as well; see Appendix A Figure 9. This is why we also performed a regression where the activism interactions were seen as a percentage of the total number of interactions, see Figure 8. This model had an R square of 0.39, a slope of 0.63, and a p-value of 0.017. The regression line indicated a positive correlation between years and the social activism interactions as a percentage of the total. Both models showed that CEO social activism is growing, which means that we found supporting evidence for hypothesis 2.

Figure 7: The linear regression of social activism over time

Figure 8: The linear regression of social activism as % of total interactions

4.5 Testing tweet vs retweet topics

To test if there is an actual statistical difference in topics between tweets and retweets we performed multiple T-tests as stated in section 3.4. The topics 'Corporate', 'Leadership', 'Politics', and 'Other'. For each parameter in the two samples, we tested the means for equality. The hypotheses with a p-value of 0.05 were:

1. Null hypothesis (H_0) : $\mu = 0$
2. Alternative hypothesis (H_A) : $\mu \neq 0$

The results of the T-tests are shown in Table 8. The null hypothesis was accepted two times and also rejected two times. This means that there is a statistically significant difference in topics between tweets and retweets. Thus, we found supporting evidence for hypothesis 3, which stated that there is a difference between CEO tweets and retweets. CEOs will for example mostly use tweets to communicate about the topic of leadership.

	Corporate	Leadership	Politics	Other
Tweet count	27758	13295	9999	36482
Retweet count	24015	4368	7680	17225
p-value	0.434635	0.00004	0.269306	0.00003
H_0 rejected	No	Yes	No	Yes

Table 8: Results of T-tests on the different parameters

5 Discussion

This study asked questions about which topics CEOs communicate on Twitter and if there are differences between tweet and retweet topics. We found that the most popular topics for CEOs were corporate, politics, and leadership. We also found that the topic of leadership is conveyed more by using tweets rather than retweets, which confirmed that these two types of interactions differ in their topics. We also found that CEOs increasingly conveyed more social activism on Twitter. In order to answer the research questions, we used an API to scrape Twitter and then transformed the scraped interactions into vectors with Sentencetransformers. The next step was to perform the unsupervised clustering with BERTopic, UMAP, and HDBSCAN. The identified clusters were then manually assigned topics based on their statistically most significant words.

5.1 Findings

Our research focused on the micro-blogging platform Twitter and its different interaction types. Using state of the art topic modelling methods, we found that CEOs mostly communicate about corporate, politics, and leadership (see Table 5). The topic corporate includes interactions that discuss company-related subjects. This can for example be about hiring new employees, thanking people, or promoting new products/services. Corporate is the largest topic identified, being more than two times the size of the second largest topic politics. This finding confirmed that CEOs mostly communicate about their own companies. The distribution of the different interactions leans heavily towards one-way communication, with replies only comprising 12,7% of the total. While multiple studies found that engaging with your audience improves your brand (Shandwick, 2012; Karaduman, 2013; Porter et al., 2015), CEOs generally still do not act on it.

The most notable finding within the topics themselves was the topic of leadership. CEOs tweet a lot about this subject, it was the second largest topic of CEO tweets (see Table 10). However, it was only the fourth largest topic for retweets (see Table 11). Our statistical test proved that this was a significant difference. This finding was in line with the theory that followers are often attracted to the leadership capabilities of CEOs (Men and Tsai, 2016). When retweeting, you are communicating indirectly with your followers by using a message/opinion from someone else. Followers want to know the personal stances and insights of CEOs as they see them as successful individuals, not from the people that they retweet. It thus makes sense that CEOs make more use of tweets instead of retweets on the topic of leadership.

Our research also contributes to a better insight into increasing CEO social activism. We found that the cluster for sustainable energy had a huge spike in activity in 2021. Compared to 2020, it saw a 480% increase (see Figure 6). In 2020, two large clusters saw a large growth, the US election cluster, and the COVID cluster (see Figure 6). The election cluster is very interesting because this amount of activity did not happen in the elections before. There always was an increase in cluster size during election years but this growth in activity is unprecedented. While the spike for the COVID cluster was expected, it also represents social activism. COVID was a large issue, and many people disagreed about which countermeasures to take. These three large positive spikes in activity, combined with the individual topics of these clusters, suggest an overall increase in social activism in 2020 and 2021 compared to the previous years. We also performed three statistical tests

to confirm that social activism is growing. This finding indicates that CEO online behavior and strategy are becoming more socially active.

5.2 Future work

A suggestion for future work would be to extract public metrics for each topic. These can then be compared to see which topics are well received by Twitter users. To make an accurate comparison, the different following sizes of the CEOs should be taken into account. A method would be to divide the public metrics of a tweet by the number of followers. The topic modeling methods described in this research could be utilized as a foundation for these future works. A better understanding of the impact of certain topics could improve social media strategies for executives.

Another research could focus more on how the clusters that represent social activism evolved over time. We identified the US election and sustainable energy clusters and their spikes, but there could be more clusters that we did not find. A deeper look at how different small topics within social activism grow and shrink could be very interesting. It would give a good insight into the evolution of CEO social media strategy.

5.3 Limitations

We used a sentence transformer model that is only trained on English data. However, in our data set were some interactions that were in another language. As a result, these interactions could not be mapped to the right position in the vector space. For example, a cluster in which the most important words were "el" and "la" was formed. Since only a small amount of interactions were in another language, it had little effect on our results. However, our results would have been more sound if we had used a multilingual model.

Another limitation is that we used all interactions that were ever made by the CEOs in our analysis. As a consequence, some interactions might have been made when they were not yet CEOs. These interactions could potentially skew our results. A solution would be to only scrape the interactions that were made when they were in fact CEOs.

A lot of interactions were not assigned a cluster by the algorithm because we used HDBSCAN, which focuses on cluster accuracy. While this makes the results gathered more robust, it also gives rise to the question if not more interactions should have been clustered. Perhaps by using a less conservative clustering algorithm, other or larger clusters would have been formed.

5.4 Conclusions

This paper looked at what CEOs share on Twitter and whether there was a difference in topics between CEO tweets and retweets. Our results found that CEOs mainly talk about the topics of corporate, politics, and leadership. We also found that, on the topic of leadership, CEOs prefer to communicate via tweets instead of retweets. Lastly, we confirmed that CEO social activism shows a growing trend. This study and its methods can be used as a foundation for further research into CEO Twitter communication.

References

- Alghawi, I. A., Yan, J., & Wei, C. (2014). Professional or interactive: Ceos' image strategies in the microblogging context. *Computers in Human Behavior*, 41, 184{189. <https://doi.org/https://doi.org/10.1016/j.chb.2014.09.027>
- Ankerst, M., Breunig, M., Kriegel, H.-P., & Sander, J. (1999). Optics: Ordering points to identify the clustering structure. *Sigmod Record* 28, 49{60. <https://doi.org/10.1145/304182.304187>
- Ante, L. (2021). How elon musk's twitter activity moves cryptocurrency markets *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3778844>
- Bendisch, F., Larsen, G., & Trueman, M. (2013). Fame and fortune: A conceptual model of ceo brands. *European Journal of Marketing* 47. <https://doi.org/10.1108/03090561311297472>
- Bielak, D., Bonini, S. M., & Oppenheim, J. M. (2007). Ceos on strategy and social issues. *The McKinsey Quarterly*, 4, 1{8.
- Bondi, T., Burbano, V., & Dell'Acqua, F. (2022). When to talk politics in business: Theory and experimental evidence of stakeholder responses to ceo political activism.
- Cambridge University Press. (n.d.). The cambridge dictionary of philosophy [Accessed 20-June-2022]. <https://dictionary.cambridge.org/us/dictionary/english/in uencer>
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in knowledge discovery and data mining* (pp. 160{172). Springer Berlin Heidelberg.
- Capriotti, P., & Ruesja, L. (2018). How ceos use twitter: A comparative analysis of global and latin american companies *International Journal of Information Management*, 39, 242{248. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2018.01.003>
- Chatterji, A. K., & To el, M. W. (2019). Assessing the impact of ceo activism. *Organization & Environment*, 32(2), 159{185. <https://doi.org/10.1177/1086026619848144>
- Childers, C. C., Lemon, L. L., & Hoy, M. G. (2019). #Sponsored #ad: Agency perspective on in uencer marketing campaigns. *Journal of Current Issues & Research in Advertising* 40(3), 258{274. <https://doi.org/10.1080/10641734.2018.1521113>
- Elliott, Brooke, W., Grant, S., & Hodge, D. (2018). Negative news and investor trust: The role of \$ rm and #ceo twitter use. *Journal of Accounting Research* 56(5), 1483{1519. <https://doi.org/https://doi.org/10.1111/1475-679X.12217>
- Emerald. (2018). Are CEOs making the grade as in uencers on twitter? *Strategic directory*, 34(3), 1{3.

- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 26{231.
- Fischer, E., & Reuber, A. R. (2011). Social interaction via new social media: (how) can interactions on twitter affect actual thinking and behavior? *Journal of Business Venturing* 26(1), 1{18. <https://doi.org/10.1016/j.jbusvent.2010.09.002>
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*
- Hambrick, D. C., & Wowak, A. J. (2021). Ceo sociopolitical activism: A stakeholder alignment model. *Academy of Management Review* 46(1), 33{59. <https://doi.org/10.5465/amr.2018.0084>
- Heavey, C., Simsek, Z., Kyprianou, C., & Risius, M. (2020). How do strategic leaders engage with social media? a theoretical framework for research and practice. *Strategic Management Journal*, 41(8), 1490{1527. <https://doi.org/10.1002/smj.3156>
- Huang, L. V., & Yeo, T. E. D. (2018). Tweeting leaders. *Internet Research* 28(1), 123{142. <https://www.proquest.com/scholarly-journals/tweeting-leaders/docview/1994329037/se-2?accountid=13598>
- Karaduman, I. (2013). The effect of social media on personal branding efforts of top level executives [The Proceedings of 9th International Strategic Management Conference]. *Procedia - Social and Behavioral Sciences* 99, 465{473. <https://doi.org/10.1016/j.sbspro.2013.10.515>
- Kim, D., Lee, J., Lee, J., & Suh, J. (2021). Elon musk's twitter and its correlation with tesla's stock market. *International Journal of Data Science and Analysis* 7, 13. <https://doi.org/10.11648/j.ijdsa.20210701.14>
- Kim, H., Park, J., Cha, M., & Jeong, J. (2015). The effect of bad news and ceo apology of corporate on user responses in social media. *PLOS ONE*, 10(5), 1{21. <https://doi.org/10.1371/journal.pone.0126358>
- Larcker, D. F., Miles, S., Tayan, B., & Wright-Violich, K. (2018). The double-edged sword of ceo activism. *Stanford Closer Look Series* 19-5.
- Lou, C., & Yuan, S. (2019). In uencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, 19(1), 58{73. <https://doi.org/10.1080/15252019.2018.1533501>
- Malhotra, C. K., & Malhotra, A. (2016). How ceos can leverage twitter. *MIT Sloan Management Review* 57(2), 73{79. <https://www.proquest.com/scholarly-journals/how-ceos-can-leverage-twitter/docview/1753248771/se-2?accountid=13598>

- Marchetti, A., & Puranam, P. (2020). Interpreting topic models using prototypical text: From ‘telling’ to ‘showing’. *INSEAD Working Paper, 2020/49/STR*. <https://doi.org/http://dx.doi.org/10.2139/ssrn.3717437>
- McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. <https://doi.org/10.48550/ARXIV.1802.03426>
- Mellon, J., & Prosser, C. (2017). Twitter and facebook are not representative of the general population: Political attitudes and demographics of british social media users. *Research & Politics, 4*(3), 2053168017720008. <https://doi.org/10.1177/2053168017720008>
- Men, L. R., & Tsai, W.-H. S. (2016). Public engagement with ceos on social media: Motivations and relational outcomes. *Public Relations Review, 42*(5), 932–942. <https://doi.org/https://doi.org/10.1016/j.pubrev.2016.08.001>
- Osnat, C.-N. (2019). Toward a conceptual model for determining ceo brand equity. *Corporate Reputation Review, 22*(4), 121–133. <https://www.proquest.com/scholarly-journals/toward-conceptual-model-determining-ceo-brand/docview/2307251384/se-2?accountid=13598>
- Ostwal, A. (2020). Test for existence of a Trend in a Time Series. Retrieved August 2, 2022, from <https://towardsdatascience.com/test-for-existence-of-a-trend-in-a-time-series-3a44f242c329>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
- Porter, M. C., Anderson, B., & Nhotsavang, M. (2015). Anti-social media: Executive twitter “engagement” and attitudes about media credibility. *Journal of Communication Management, 19*(3), 270–287. <https://www.proquest.com/scholarly-journals/anti-social-media-executive-twitter-engagement/docview/2120772338/se-2?accountid=13598>
- Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. <https://arxiv.org/abs/1908.10084>
- Shandwick, W. (2012). The social ceo: Executives tell all [Accessed 24-May-2022]. <http://www.webershandwick.com/uploads/news/files/Social-CEO-Study.pdf>
- Twitter. (2022). Twitter api [Accessed 12-May-2022]. <https://developer.twitter.com/en/docs/twitter-api>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J.,

Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Wolfe, S. J. (2019). Business playing politics: Strengthening shareholders' rights in the age of ceo activism. *Lewis & Clark L. Rev.*, 23, 1469.

Xia, L. (2013). Effects of companies' responses to consumer criticism in social media. *International Journal of Electronic Commerce*, 17(4), 73–100. <https://doi.org/10.2753/JEC1086-4415170403>

Zohrehvand, A. (2021). Search-tweets-python [Accessed 12-May-2022]. <https://github.com/twitterdev/search-tweets-python>

Zohrehvand, A. (2022). Fifty million followers can't be wrong, or can they? effects of social media feedback on ceo communication. *SSRN Electronic Journal*. <https://doi.org/https://dx.doi.org/10.2139/ssrn.4141444>

A Count of Twitter interactions overtime

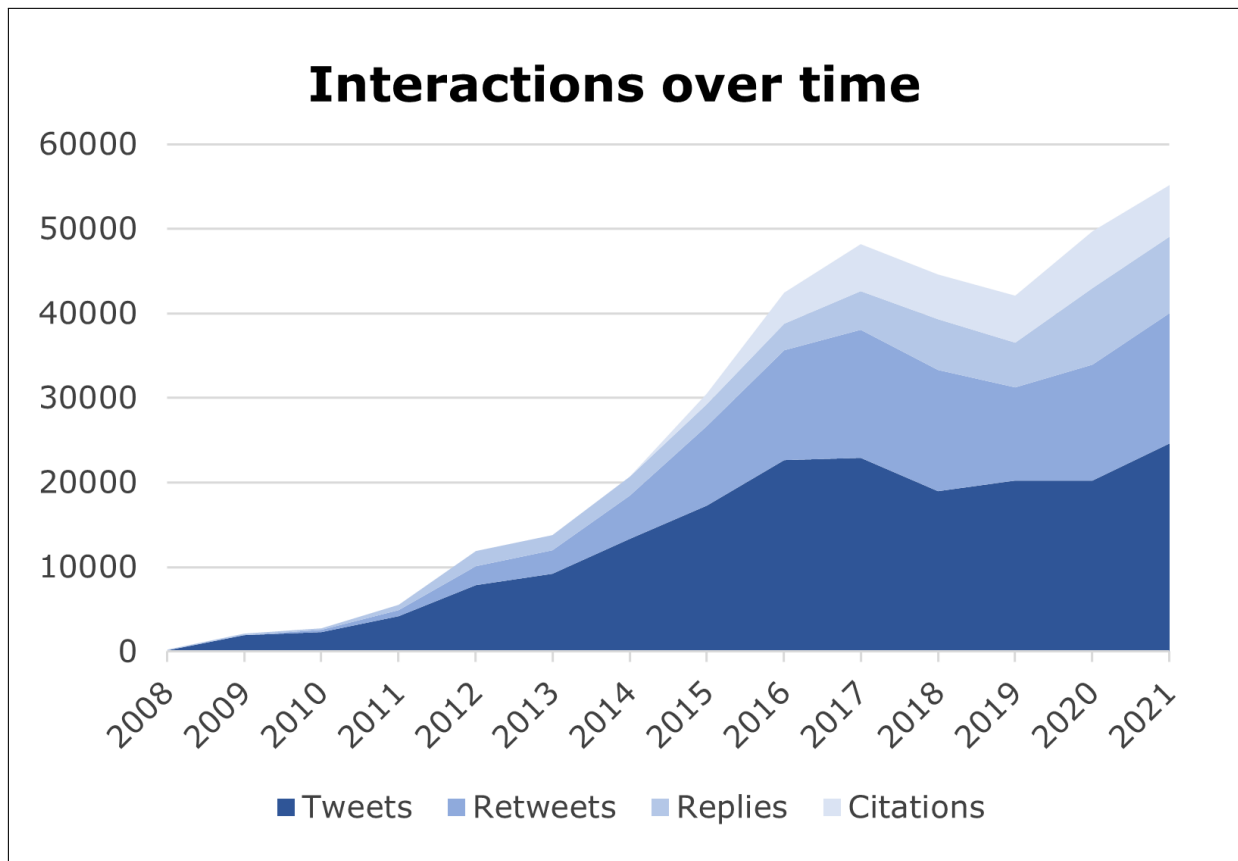


Figure 9: Twitter interactions over time

B Cluster figures

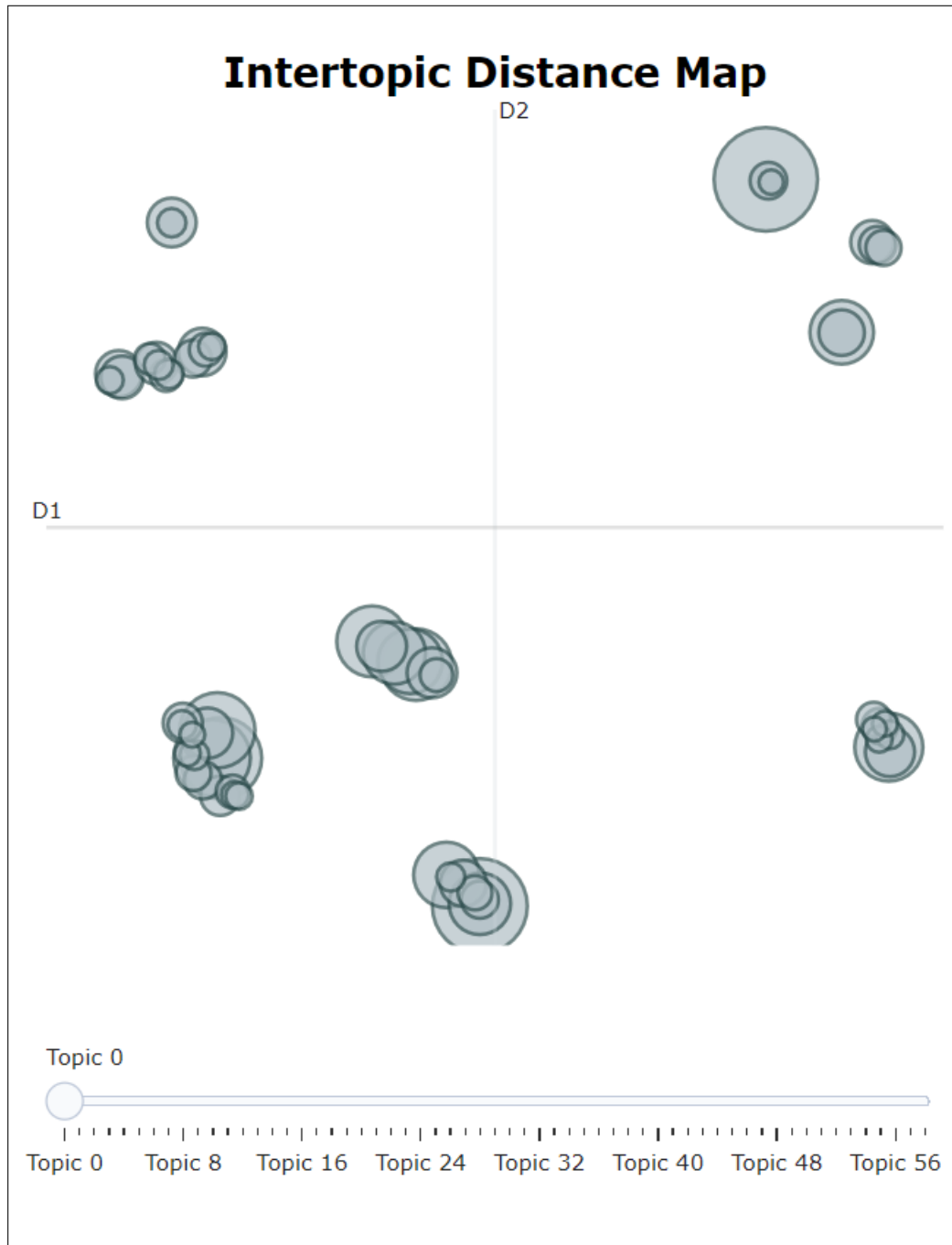


Figure 10: Clusters found in all tweets, closeness generally represents similarity

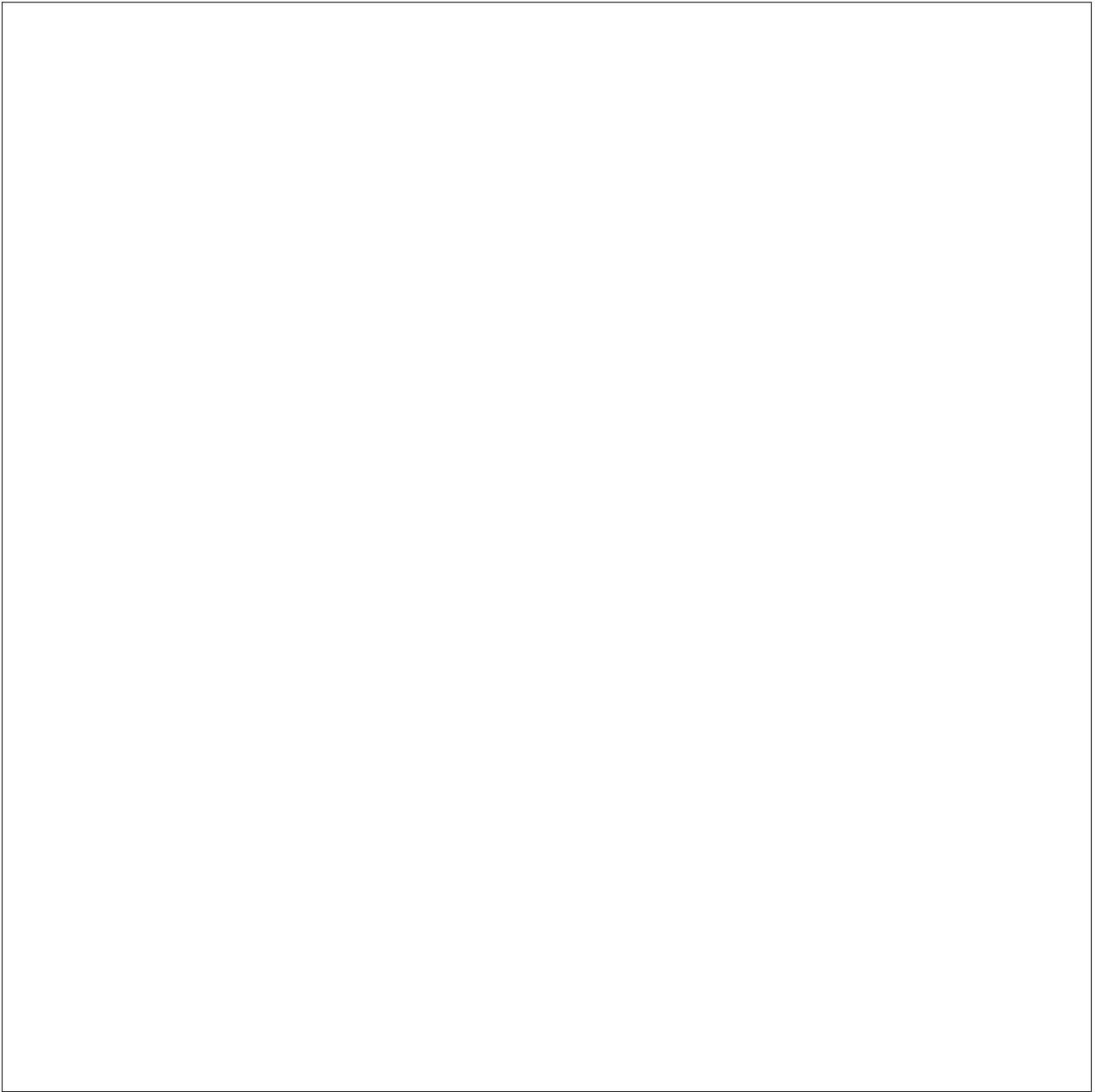


Figure 11: Cosine similarity matrix for clusters found in all tweets



Figure 12: Clusters found in retweets, closeness generally represents similarity

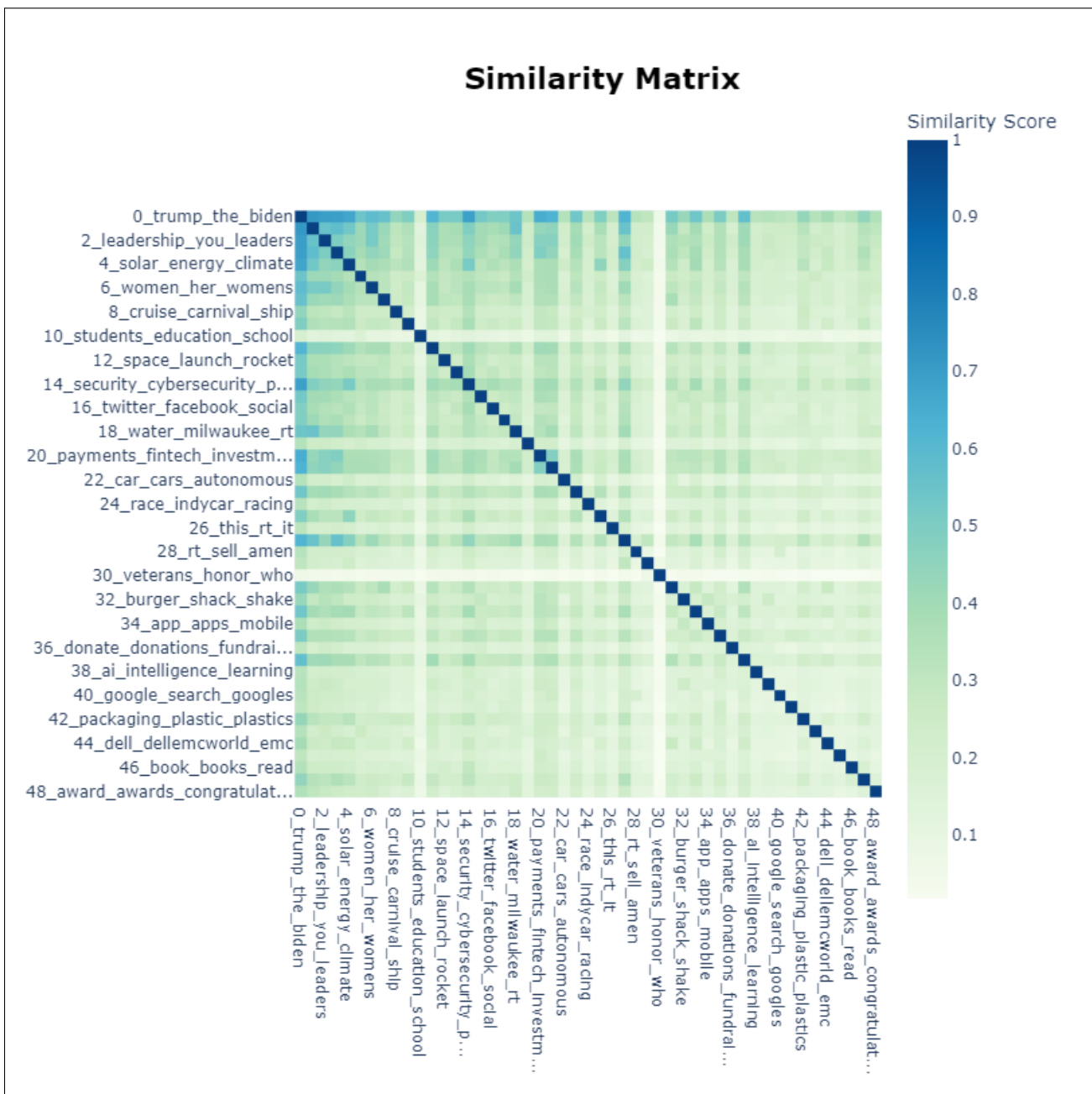


Figure 13: Cosine similarity matrix for clusters found in the retweets

C Clusters over time

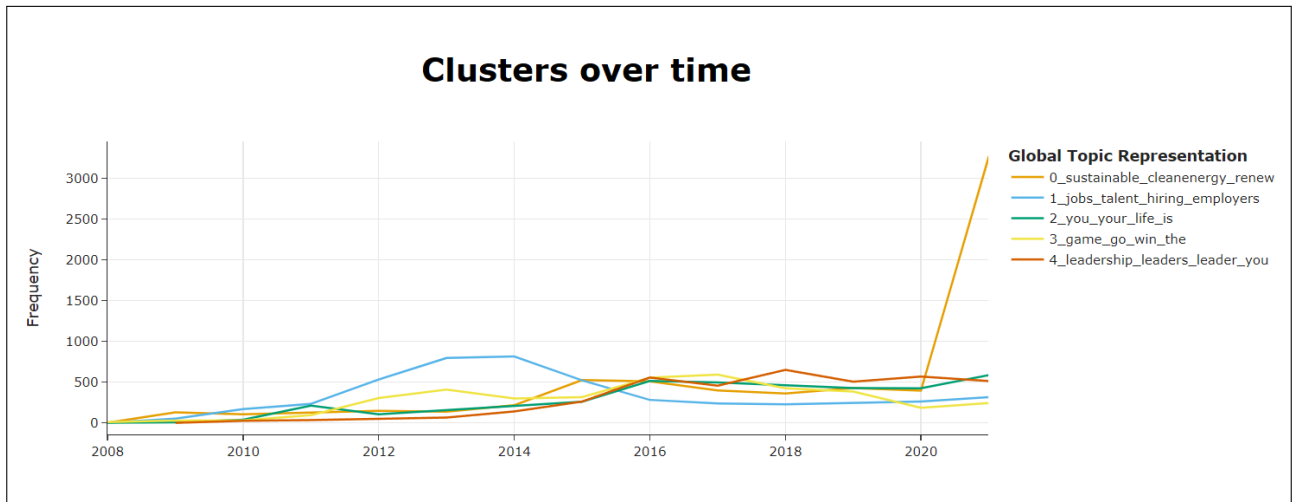


Figure 14: Largest clusters of tweets over time, 2022 data is left out

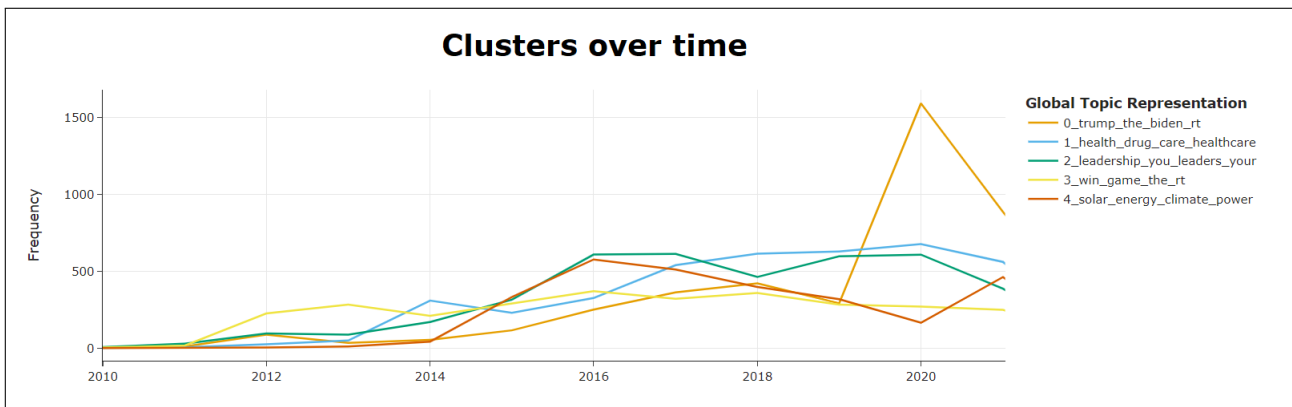


Figure 15: Largest clusters of retweets over time, 2022 data is left out

D Twitter interaction topics

Topic	Interaction count
Corporate	75379
Politics	19710
Leadership	15999
Sports	13838
Climate	11271
Covid	6636
Social media	6622
Family	4987
Cryptocurrencies	4079
Charity	3845

Table 9: Ten largest general topics of all interactions

Topic	Tweet count
Corporate	27758
Leadership	13295
Politics	9999
Climate	7577
Sports	7266
Social media	3844
Good morning	3280
Covid	2955
Music	2690
Family	2276

Table 10: Ten largest general topics of tweets

Topic	Retweet count
Corporate	24015
Politics	7680
Sports	4379
Leadership	4368
Climate	2921
Covid	2510
Music	1355
Social Media	1171
Charity	1141
Cryptocurrencies	1099

Table 11: Ten largest general topics of retweets