Temporal analysis of topic modeling output by machine learning techniques

Faezeh Azizi[1] · Hamed Vahdat-Nejad[1] · Hamideh Hajiabadi[2]

Abstract

Topic modeling is widely recognized as one of the most effective and significant methods of unsupervised text analysis. This method facilitates identifying and extracting topics in document sets associated with various entities (e.g., countries, websites, journals, etc.). Nonetheless, the method's output lacks high-level information per entity. Applying machine learning methods to topic modeling outputs is generally challenging. Some studies have already applied machine learning methods statically, ignoring the effect of time on topic modeling outputs. The inclusion of time introduces additional complexity to the problem. This study introduces a novel approach to clustering the output of topic modeling per entity, considering the time factor. Topic popularity over time and the feature vector for each entity over time are proposed for this purpose. Due to the high dimensionality of the proposed feature vector, selecting an appropriate dimension reduction technique and the corresponding clustering algorithm may not be a straightforward task. This research proposes a new approach to selecting a dimensionality reduction method and its corresponding clustering technique. A case study is conducted on COVID-19-related tweets to evaluate the proposed method's performance. The proposed approach applies t-distributed stochastic neighbor embedding for dimensionality reduction and fuzzy C-means for clustering. While our study incorporates the time factor, unlike previous research, it also outperforms them in terms of the Davies–Bouldin index, silhouette coefficient, Calinski–Harabasz index, and Dunn index parameters. The proposed method enables researchers in natural language processing to analyze topic dynamics across various entities, leading to improved research outcomes.

Keywords Topic modeling · Natural language processing · Clustering · Social network · COVID-19

1 Introduction

Topic modeling [1] is widely recognized as a highly effective and significant technique for unsupervised text analysis. This method enables the identification of topics in document sets and their extraction from various texts, such as research article abstracts [2] and social network user comments [3]. In topic modeling, words' weights are extracted to identify significant and frequently occurring words. Nonetheless, these weights alone do not offer high-level information and necessitate additional analysis and processing through machine learning techniques. Applying machine learning techniques to the output of topic modeling can be challenging, especially when dealing with big textual data that represent different entities over time (e.g., countries, journals, and authors).

Formally, let's consider a scenario where we have a collection of X texts (e.g., tweets, paragraphs, or documents) related to Y entities (e.g., countries, websites, or journals). These texts are associated with various time slots. Topic modeling, as a method, statically provides the topics of the texts and their associated word probabilities as output. However, it does not account for the time factor or offer additional insights about the entities themselves. This leads to the crucial question of extracting higher-level information from the outputs of topic modeling. In other words, topic modeling

✉ Hamed Vahdat-Nejad
vahdatnejad@birjand.ac.ir

Faezeh Azizi
faezeh.azizi1995@birjand.ac.ir

Hamideh Hajiabadi
hamideh.hajiabadi@kit.edu

1 Perlab, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran

2 Institute for Program Structures and Data Organization, Karlsruhe Institute of Technology, Karlsruhe, Germany

alone does not analyze the entities directly over time. Consequently, another question arises: How can we effectively cluster the entities using the topic modeling output?

Previous research has presented static analyses on topic modeling output. In this context, Euclidean methods [4] and Jensen–Shannon divergence (JSD) [2, 5] have been used to cluster entities based on their static distributions of topics. However, the primary challenge arises when considering the element of time. Indeed, the static clustering of entities, without taking into account the time factor, will not yield accurate results. In numerous applications, the topics of entities undergo changes over time. Figure 1 presents the temporal changes in certain topics related to COVID-19 pandemic in the United States of America (USA). Specifically, the topics of Vaccine, Masking, and Work and telecommuting are highlighted. Incorporating the element of time will increase the complexity of the problem. To the best of our knowledge, none of the previous research studies have taken the time factor into consideration [2, 4, 5].

Continuing previous work [3], this study proposes a new method for clustering the output of topic modeling over time and per entity. Given that one of the largest, most significant, and most widely-discussed text datasets in recent years pertains to COVID-19, our topic modeling analysis focuses on a dataset of tweets about COVID-19 for 2020 and 2021. To consider the time factor, we introduce a parameter of the popularity of topics over time. This parameter is then utilized to propose a temporal feature vector for each entity (country). Given the large dimensions of the proposed feature vector, it is necessary to select a dimension reduction technique for big datasets. Moreover, an optimal clustering method must be selected. Selecting the appropriate combination of dimension reduction and clustering methods can be a daunting task. This study proposes an approach to selecting a pair of dimension reduction and clustering methods based on DBI [6], SC [7], CHI [8], and DI [9] clustering evaluation criteria.

After applying the proposed method, the t-SNE [10] technique was selected for dimension reduction, and the FCM [11] algorithm was chosen for clustering. While previous research overlooked the time factor, the results of the proposed method show promising performance in comparison to previous methods, as indicated by favorable evaluation parameters, including DBI, SC, CHI, and DI. In summary, the main contributions of this research are outlined as follows:

- Proposing a temporal clustering scheme of the output of topic modeling.
- Proposing a pair selection method for dimension reduction and clustering.

These outcomes offer valuable insights for natural language processing researchers, enabling them to better understand the dynamics of topic analysis for various entities and potentially improve the accuracy of their research findings.

The article is organized as follows. Section 2 provides a comprehensive review of the research background. The methodological approach is detailed in Sect. 3. Section 4 presents the research experiments, including comparing the proposed method with previous methods. Finally, Sect. 5 provides a summary and conclusion.

2 Related work

Words are arranged in accordance with syntactic patterns to form sentences, and these sentences collectively constitute a text in relation to one another. When considering all the sentences in a text, one encounters a complex communication structure and a collection of explicit and implicit relationship concepts [12]. Topic modeling is a widely used method for discovering implicit relationships in texts [3].

Autoencoders are a type of the topic modeling method that aim to learn a low-dimensional representation of the input data by encoding it into a latent space. This latent space is then used to extract topics or themes present in the text. Autoencoders work by capturing the underlying patterns in the data through neural networks or other machine learning algorithms. Variational autoencoders (VAEs) [13] are one of the recent popular algorithms for topic modeling [14]. This encoder has been used in various fields, including emotion classification [15, 16], sentiment classification [17], information extraction enhancement [18], and conversational semantic role labeling enhancement [19]. With the help of VAE, various goals such as reducing the computational cost of computing the posterior distribution [14], improving multi-label emotion classification [15, 16], improving sentiment classification at the document level [17], and improving information extraction [18] have been realized.

On the other hand, latent Dirichlet allocation (LDA) is a popular probabilistic generative model for topic modeling [3]. LDA is based on the assumption that each document composed of a combination of a group of topics, and each word in the document is generated from one of these topics. By inferring the underlying topic structure from the observed word distributions, LDA is able to identify the topics present in a given text corpus [3]. This approach has been successfully applied in diverse fields, such as geography [20], psychology [3], news [21], and economics [22]. Through topic modeling, several objectives have been achieved with minimal human intervention. These include the identification of current and

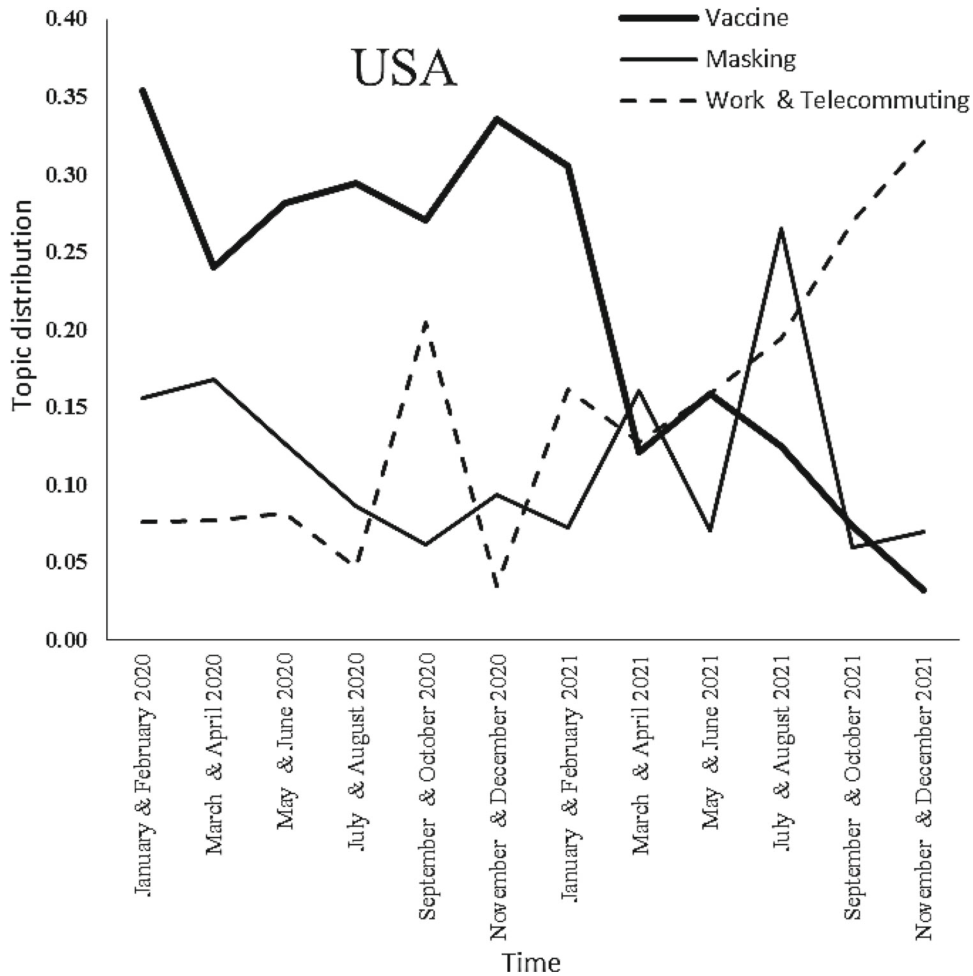


Fig. 1 Distribution of three topics of the USA over time

prevalent research topics [23], enhancement of search capabilities within the portal of the Federal Digital Libraries of the USA [24], and the development of an ontology graph [25].

While both autoencoders and LDA can be used for topic modeling, LDA has been shown to have several advantages over autoencoders. One key advantage of LDA is its interpretability. Since LDA is based on a probabilistic generative model, the resulting topics are easily interpretable as distributions over words. This allows users to understand and analyze the topics discovered by LDA more easily compared to the latent representations learned by autoencoders. Additionally, LDA is robust to noise and sparse data, making it well-suited for working with real-world text data that often contains noise and missing information. Its probabilistic framework also provides a principled way to model uncertainty and capture the inherent variability in language data. Overall, while both autoencoders and LDA can be effective for topic modeling, LDA's interpretability, robustness, and probabilistic nature make it a superior choice for extracting meaningful and interpretable topics from text data.

After all, analyzing topic modeling outputs, particularly through clustering, is a new research area. Since topic modeling is unsupervised and lacks labeled data, clustering becomes a crucial analysis. Only a few studies have explored topic modeling results using clustering, which will be discussed in detail below.

To cluster countries based on their article publication patterns in the field of environmental science, researchers gathered a dataset of more than 3000 articles published from 2005 to 2019 from relevant journals [4]. The latent Dirichlet allocation (LDA) topic modeling method [26] was utilized to analyze the abstracts of these articles, leading to the identification of 20 topics, such as environmental impact assessment and improved clean cookstoves. The topic ratios' distributions were statically calculated using LDA for 17 countries, and Euclidean distances were employed to measure the dissimilarities between these distributions. Finally, hierarchical clustering was applied to cluster the countries based on these distances.

Similarly, for clustering countries based on their publishing patterns in the field of structural engineering, researchers gathered a dataset comprising over 51,000 articles published from 2000 to 2020 from relevant journals [5]. The LDA topic modeling method was applied to the abstracts, resulting in the identification of 50 topics, such as structural control, wind flow, and turbulence. The distributions of these topics across 31 countries were calculated statically using LDA, and the Jensen–Shannon divergence (JSD) parameter was used to determine the distances between the topic distributions. Ultimately, hierarchical clustering was employed to cluster the countries based on these distances.

Finally, more than 17,000 articles published in journals in the field of transportation between 1990 and 2015 have been compiled for the purpose of clustering countries based on their article publication patterns [2]. The LDA algorithm was used to analyze the abstracts of these articles, resulting in the identification of 50 topics, such as travel behavior and non-motorized mobility. Next, the distributions of topic ratios in 32 countries were determined using LDA. Additionally, the distances between these distributions were calculated using JSD. Finally, the distances were clustered using the hierarchical clustering method.

None of the above studies have included the time factor when analyzing the outputs of topic modeling for entity clustering. Indeed, the entity of countries in previous research is statically clustered. The clustering process has ignored the fact that topics change over time. The static analysis of the datasets results in the loss of a significant portion of the information (Fig. 1).

3 Proposed clustering method

The topic modeling method proposed in previous research [3] was initially employed to extract relevant topics. Following that, the concept of topic popularity is introduced, and based on that, the feature vector derived from the output of topic modeling over time is proposed for countries. Next, the proposed approach to select the suitable method for dimension reduction, clustering, and cluster number determination is introduced. Ultimately, the clustering results are presented. The outline of the proposed method is provided in Fig. 2.

A case study is conducted on a dataset containing 14 million tweets (in 2020 and 2021) related to COVID-19. Then, 120,000 tweets are randomly selected for every 2-month period to normalize the dataset. Ultimately, the dataset contains a total of 1,440,000 tweets for 12 2-month periods in 2020 and 2021. After applying topic modeling [3], the word cloud of each cluster is utilized to determine each topic's title [5].To this end, larger words carry more significance, and the selection of each cluster's name is based on the larger words in the word cloud. Below are the keywords and the chosen name for each topic.

Masking = {mask, wear, gloves, protects, facemasks, spread, face}
Cases and death cases = {confirm, case, death, passed, highest}
Spread = {variant, global, spread, outbreak, infect}
Work and telecommuting = {office, work, home, workplace, employ}
Students and education = {children, university, student, college, exam}
Vaccine = {scientist, passport, vaccine, immunity}
Power and politics = {Biden, president, Trump, government}
Economy = {business, economy, market, work}
Voluntary affairs = {service, humanity, warrior, volunteer, member}
Plasma = {plasma, therapy, donors, recovery, immunity, blood}

3.1 Popularity of topic

After topic extraction, topic popularity is proposed as a means to extract features from the results of topic modeling over time. This concept demonstrates the weight of a topic in relation to other topics in a specific period, achieved through the analysis of topic distributions over time. In other words, the popularity of topic K in the t^{th} 2-month period depends on the number of tweets that contain this topic during this period to the total number of tweets in this period. Therefore, by utilizing the concept of topic popularity, we can easily compute and analyze the impact and significance of topic K during time t. The popularity of topic K in the t^{th} 2-month period ($p_{\text{K}}(t)$) is defined as the ratio of the number of tweets related to topic K to the total number of tweets in this period:

$$(p_K(t)) = \frac{\text{Tweets related to topic } K \text{ in the } t^{\text{th}} \text{ 2\_month period}}{\text{Tweets in the } t^{\text{th}} \text{ 2\_month period}} \tag{1}$$

The popularity of topic K throughout the entire period demonstrates the total extent to which topic K has been addressed and how popular it has been among users. P_K, the popularity of topic K over the entire period, is defined as follows:

$$P_K = \sum_{t=1}^{12} p_K(t) \tag{2}$$

3.2 Proposed feature vector

In order to apply the machine learning techniques to the topic modeling results, it is necessary to model countries as feature vectors that consider the time trend. If the objective of using machine learning methods is to perform a static analysis of

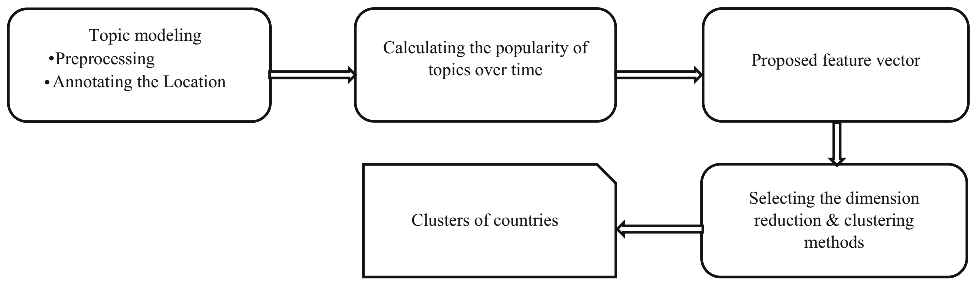


Fig. 2 The outline of the proposed method

topics over the entire period, P_K would be sufficient for any topic, as it takes the entire period into account. However, since the time factor is particularly important in most research studies, the feature vector consists of a sequence of p_K(t)s. Indeed, a feature vector of topic popularity values over time intervals is proposed for each country. In this case study, the feature vector is formed with a length of 120 for this purpose:

$$\text{Feature vector of a country} = \{\underbrace{p_1(1), p_1(2), \ldots, p_1(12)}_{\text{The popularity of the first topic in subsequent intervals}}, \ldots, \underbrace{p_{10}(1), p_{10}(2), \ldots, p_{10}(12)}_{\text{The popularity of the last topic in subsequent intervals}}\} \tag{3}$$

According to the proposed Eq. (3), the feature vector of countries is derived from the set of the ten topics' popularities during the 12 2-month periods. Next, the modeled data set is accumulated as a matrix in which each country represents a row. In this case study, the dataset contains the feature vectors of 32 countries, and its dimensions are 32*120, as below:

$$\text{Dataset} = \begin{bmatrix} \text{Feature vector of the first country} \\ \text{Feature vector of the second country} \\ \ldots \\ \text{Feature vector of the 32nd country} \end{bmatrix}_{32*120} \tag{4}$$

3.3 Clustering

In order to apply machine learning methods to big high-dimensional datasets, we need to perform dimensionality reduction. In high-dimensional datasets, the presence of numerous metadata and features per data item can make it difficult and time-consuming to identify the appropriate features. Furthermore, the computational complexity in high dimensions presents a significant challenge in training machine learning methods, particularly for traditional approaches. This complexity may lead to time-consuming and difficult training processes, potentially hindering the proper functioning of machine learning methods. Consequently, these methods may exhibit a high error rate and low accuracy.

As numerous techniques have been proposed for dimensionality reduction, our aim is to identify the most effective methods for mapping high-dimensional datasets to lower dimensions in this subject. Specifically, we seek suitable dimensionality reduction approaches for datasets that exhibit significant scattering in higher dimensions.

To address this, we utilize two nonlinear dimensionality reduction methods: t-SNE [10] and UMAP [27]. These techniques excel in handling sparsely distributed data in high-dimensional spaces [28]. They achieve dimensionality reduction by projecting the data into a new space while preserving the relative distances between the data points [28]. In essence, they provide a compressed representation of the high-dimensional space while maintaining the proximity and distances between the data points. t-SNE and UMAP are commonly employed in textual data analysis and similar applications for dimension reduction [28].

Besides, since the data in this study is unlabeled, clustering is employed to group objects in the dataset and form meaningful clusters [29]. Multiple clustering algorithms exist, each with unique properties and characteristics, aiming to ensure similarity within clusters and distinctiveness between them. This case study focuses on three traditional clustering algorithms: K-means [30], FCM [11], and hierarchical clustering (HC) [31], which are among the most widely used and effective clustering algorithms [32]. K-means partitions data into clusters based on similarity and iteratively optimize the cluster center weights. K-means is a simple and efficient method widely used in various domains due to its scalability, interpretability, and capability to produce clear clusters. FCM, a fuzzy variant of K-Means, is effective for clustering noisy data by assigning data points to clusters based on their weights and iteratively optimizing the cluster center weights. FCM offers soft clustering, accommodating multiple cluster memberships, and noise robustness [32]. HC utilizes hierarchical clustering to divide data into smaller clusters, recursively, generating the entire cluster hierarchy. HC generates a hierarchical structure, allowing for cluster interpretation without prior knowledge of cluster numbers. It provides visual representations of cluster relationships and similarities at different levels [33].

The study applies these three clustering methods to countries with and without reducing the dataset's dimensions. It enables obtaining high-level and analyzable information, facilitating the comparison among similar countries. By employing two dimension-reducing methods (as well as one none-reducing method) and three clustering methods, the study generates nine model pairs. The determination of the best methods will be discussed later in the paper.

3.4 Selecting the dimension reduction and clustering pair

In order to select the optimal dimension reduction and clustering methods, we employ four key clustering criteria: Davies–Bouldin index (DBI) [6], silhouette coefficient (SC) [7], Calinski–Harabasz index (CHI) [8], and Dunn index [9]. These criteria are widely recognized for their effectiveness in revealing the underlying clustering structure within a dataset [33]. By evaluating them, we can determine the most suitable dimension reduction and clustering techniques for our study.

Davies–Bouldin index [6] is a clustering evaluation metric used to assess the quality of data partitioning into clusters. The calculation involves considering the distance between cluster centers and the internal variance of the clusters. This measure computes the distance between the center of each cluster and the centers of other clusters. It then divides this distance by the sum of the internal variance of that cluster. Subsequently, the ratio is computed for each cluster. This measure is equivalent to the mean of these ratios. It quantifies the average maximum ratio of dispersion within clusters to dispersion between clusters. It evaluates the output value for the number of clusters and identifies the clustering algorithm that maximizes inter-cluster distance and minimizes intra-cluster distance. A lower DBI value indicates better clustering performance. Dunn index (DI) [9] is a clustering evaluation criterion that quantifies the clustering quality. This measure assesses clustering quality by quantifying the distance between points within a cluster and points between clusters. It is calculated by dividing the smallest distance between any two cluster centroids (known as the lowest inter-cluster distance) by the largest distance between any two points within any cluster (known as the highest intra-cluster distance). A higher DI value indicates better clustering performance. The silhouette coefficient (SC) [7] is a clustering evaluation criterion that assesses clustering quality by quantifying the distances between points within a cluster and the distances between points in neighboring clusters. It is calculated as the mean difference between the distance of a point from all points within its own cluster and the distance from all points in neighboring clusters, divided by the larger of these two distances. A higher coefficient of this measure indicates superior performance. Finally, Calinski–Harabasz index (CHI) [8] is based on the ratio of the inter-cluster variance to intra-cluster variance. The clustering quality is better when the inter-cluster variance is larger and the intra-cluster variance is smaller. Moreover, the clustering is superior, the higher the CHI index value. Based on the preceding subsection, nine established cases are available for selecting the dimensionality reduction and clustering methods. However, determining the most suitable methods poses a challenge. Two approaches, visual and numerical, are proposed to aid in selecting the appropriate methods. To this end, the evaluation criteria, including CHI, SC, DBI, and DI, are calculated for all cases. Then, each criterion's lowest and highest values across all cases determine the interval for that measure. Next, each of the four criteria's intervals is divided into four equal subintervals by introducing three thresholds. In Table 1, the threshold values are denoted as Good, Normal, and Bad. The interval is specified by best and worst values.

The visual method uses four grayscale colors for the four subintervals of each criterion, with the best mode being the darker color and the worst state being the lighter color. Further discussion is provided in the next section.

In the numerical method, the criteria's values are scaled between zero and one hundred to facilitate an accurate examination. The scaled value is calculated as follows:

$$X_{new} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \quad (5)$$

$x_{\min}$ is the lowest interval value, $x_{\max}$ is the largest value of the considered evaluation criterion, and the symbol x represents the value of the criterion. This normalization is accomplished by ensuring that the best value of an interval is represented as 100 while the worst value is 0. As for the DBI criterion, a lower value is preferable; its scaled value is subtracted from one hundred. Afterward, the criteria's normalized values of each feature reduction and clustering pair are averaged. Finally, by comparing the averaged values of different methods, the optimal clustering, dimension reduction, as well as cluster number will be determined.

Table 1 Threshold values for evaluation criteria

Threshold					
Best	Good	Normal	Bad	Worst	Metric
26.7434	20.1787	13.6141	7.0495	0.4849	CHI
0.3443	− 0.0459	− 0.4361	− 0.8263	− 1.2165	SC
0.7357	2.4933	4.2509	6.0084	7.7660	DBI
0.00058	0.00046	0.00035	0.00023	0.00012	DI

4 Experiment

The main part of the implementations has been done in the Python programming environment on Windows 10, 64-bit. The hardware was a computer with corei5, 3.4 GHz frequency, and 6 megabytes of level three (L3) cache memory processor, 4 GB of RAM, and 2 GB of graphics card. The dataset used includes 14 million tweets related to the Covid-19 virus from the years 2020 and 2021. In order to normalize, 120,000 tweets have been randomly selected for each two-month period. Finally, the dataset containing 1,440,000 tweets for 12 2-month time periods has been used. To tag the location of tweets related to each country, the common natural language processing framework GATE has been used [3]. This framework includes various tools that are used to process and analyze text, identify and categorize information, extract information, and produce various outputs such as reports and charts. In order to preprocess tweets, the required functions were called from the Gensim library, which is one of the most famous libraries used in topic modeling [3].

At first, we present the experiments for extracting static topics without considering the time factor. Subsequently, the topics are dynamically evaluated, taking into account the temporal aspect. Finally, the proposed method is compared with previous works. The case study is performed for countries that have sufficient tweets in the dataset.

4.1 Static topics

Topics are initially investigated regardless of time. Table 2 shows the three hottest topics for each country. The vaccine has been the leading topic for nearly all countries, as indicated in the table. Work & telecommuting has emerged as the second topic in most countries. The top three topics exhibit a consistent trend across most of the countries, indicating a similar tendency among them. Figure 3 illustrates the worldwide distribution of topics based on the global percentage of tweets. Among these topics, vaccine exhibits the highest percentage, while Work and telecommuting and Masking rank second and third, respectively.

4.2 Dynamic topics

Table 2 indicates that the USA and South Korea have exhibited similar behavior. However, when we look at the details (i.e., the trend of the topics as shown in Fig. 4), we can see that their topics exhibit dissimilar trends. In Fig. 4, the trends of these topics vary significantly between the two countries. As such, a dynamic analysis of topics is required. This subsection examines the influence of the time factor. The popularity of topics ($p_k(t)$) values for the entire world are presented in Table 3. Vaccine, Work and telecommuting, and Masking are the most popular topics on a global scale. Figure 5 illustrates the popularity trend of these topics worldwide.

Tables 4, 5, and 6 show the popularity of topics in the three countries with the highest tweet volume, including the USA, India, and China. The topics in topic popularity tables are arranged in descending order based on their P_k values. Appendix 1 presents the topic popularity tables for other countries, including Australia, the UK, Canada, Pakistan, Japan, Germany, France, Ireland, Singapore, UAE, Mexico, Italy, Sweden, Brazil, Iran, Russia, South Korea, Spain, Switzerland, Turkey, Netherlands, Belgium, Denmark, Chile, Saudi Arabia, Peru, Portugal, and Qatar.

Figure 6 illustrates the trend of hot topics in the USA, India, and China. Appendix 2 shows the trend of hot topics for other countries, including Australia, United Kingdom (UK), Canada, Pakistan, Japan, Germany, France, Ireland, Singapore, UAE, Mexico, Italy, Sweden, Brazil, Iran, Russia, South Korea, Spain, Switzerland, Turkey, Netherlands, Belgium, Denmark, Chile, Saudi Arabia, Peru, Portugal, and Qatar.

4.3 Evaluation

Figure 7 compares the performances of the three clustering methods K-means, FCM, and HC, before and after dimension reduction using UMAP and t-SNE in terms of the discussed clustering criteria (CHI, SC, DBI, and DI). Besides, seven different cluster numbers (3–9 clusters) are evaluated. Each row in the table corresponds to one of these criteria, while each column represents a specific cluster number (k) value. By visually inspecting, the FCM clustering and the UMAP

Table 2 Hottest topics per country. Numbers 1, 2, and 3 represent the first, second, and third hot topics, respectively

Country	Topic									
	Spread	Work and telecommuting	Students and education	Vaccine	Power and politics	Economy	Voluntary affairs	Plasma	Cases and death cases	Masking
USA		3		1						2
India	3			1					2	
China				1	2		3			
Australia		2		1					3	
UK		2		1	3					
Canada		2		1						3
Pakistan				1	3				2	
Japan				1		3			2	
Germany				1					2	3
France		2		1					3	
Ireland		2		1					3	
Singapore		2		1						3
UAE		2		1					3	
Mexico		3		1					2	
Italy		2		1					3	
Sweden				1		3			2	
Brazil		3		2					1	
Iran	2			3					1	
Russia		3		1				2		
South Korea		3		1						2
Spain		3		1					2	
Switzerland		2		1						3
Turkey		3		1					2	
Netherlands		2		1	3					
Belgium		2		1					3	
Denmark				1	3					2
Chile		3		1					2	
Saudi Arabia	3	2		1						
Peru		3		1					2	
Portugal		2		1					3	
Qatar		2		1	3					2
Ecuador		2		1					3	

dimension reduction methods (Fig. 7b) result in the darkest colors.

After selecting the FCM clustering and UMAP dimension reduction methods, we look for the darkest column to determine the number of clusters. The leftmost column, i.e., three clusters, is clearly the darkest and therefore selected. It is also the darkest column in the figure. Table 7 presents the numerical verification of the values following the visual approach. It contains rows with CHI, SC, DBI, and DI evaluation criteria values, which have been normalized using Eq. (5). The average of normalized criteria for each state is provided at the end of the columns. Numerical analysis confirms that the FCM clustering method achieves its highest average value for three clusters when combined with the UMAP dimension reduction technique. The clustering results are as follows.

Fig. 3 Percentage of topics in the dataset

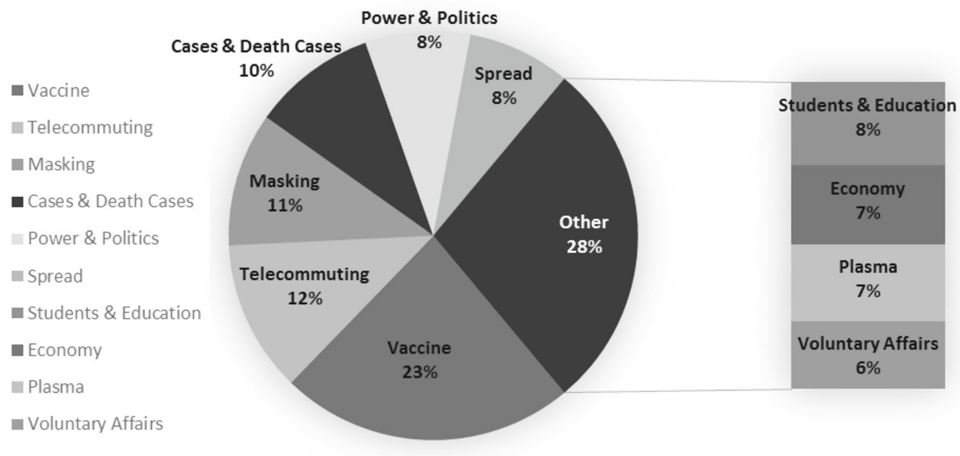
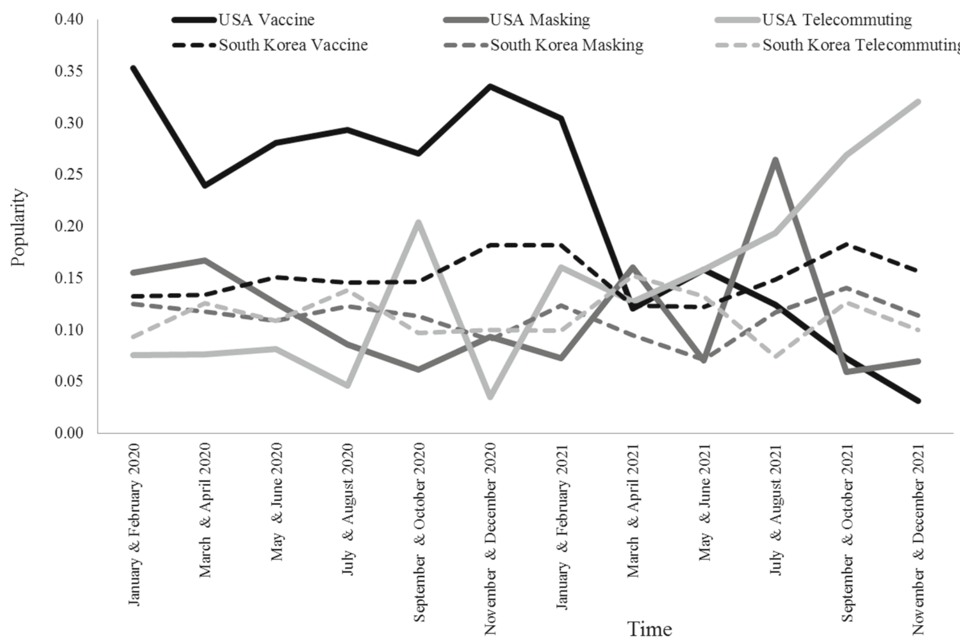


Fig. 4 Topics popularity trends in South Korea and the USA



- Russia, Netherlands, France, Spain, Switzerland, UK, Peru, and Portugal.
- USA, Australia, Belgium, Canada, Pakistan, Japan, Germany, Ireland, UAE, Italy, Denmark, Chile, and Qatar.
- India, China, Singapore, Mexico, Sweden, Brazil, Iran, South Korea, Turkey, and Saudi Arabia.

This research focuses on analyzing the output of topic modeling through clustering. Previously, Palanichamy et al. [4], Sun and Yin [2], and Xie et al. [5] have conducted static analyses on the output of topic modeling. They have utilized static clustering of topics using the Euclidean [4] and JSD [2, 5] methods. Sun and Yin [2] and Xie et al. [5] have employed a similar method. A comparison is made in Table 8 in terms of the evaluation criteria, including CHI, SC, DBI, and DI. It should be noted that higher values of CHI, SC, and DI and lower values of DBI are preferable. Consequently, in all criteria, the proposed method has achieved the best performance among all available methods.

Table 3 The popularity of topics in the world

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.3049	0.3048	0.2905	0.3531	0.1676	0.125
Telecommuting	0.038	0.0499	0.094	0.1141	0.2003	0.2149
Masking	0.0958	0.1087	0.1333	0.0862	0.1094	0.0805
Students	0.0109	0.0349	0.0295	0.03894	0.0807	0.1137
Death	0.1323	0.1232	0.1152	0.1305	0.1083	0.0916
Voluntarily	0.0515	0.0427	0.0378	0.0532	0.053	0.0202
Plasma	0.0803	0.0036	0.084	0.0063	0.0503	0.1098
Economy	0.0437	0.0739	0.0647	0.0708	0.0898	0.093
Spread	0.1236	0.1324	0.0644	0.0587	0.0718	0.08
Politics	0.1189	0.1259	0.0865	0.0879	0.0687	0.0713

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.0894	0.0935	0.1365	0.1961	0.1549	0.3813	2.5976
Telecommuting	0.1167	0.1319	0.1544	0.1427	0.1961	0.1811	1.6343
Masking	0.1263	0.1037	0.0981	0.1125	0.1269	0.1183	1.2998
Students	0.1818	0.168	0.1481	0.2107	0.0414	0.0144	1.0731
Death	0.0882	0.0623	0.0264	0.039	0.0445	0.0129	0.9744
Voluntarily	0.0827	0.0566	0.1019	0.0894	0.1944	0.1748	0.9584
Plasma	0.0454	0.1985	0.1666	0.0661	0.0718	0.0119	0.8946
Economy	0.095	0.0621	0.062	0.0639	0.0987	0.0632	0.8807
Spread	0.0772	0.0615	0.0688	0.0506	0.0369	0.0239	0.8498
Politics	0.0972	0.0619	0.0372	0.0291	0.0345	0.0181	0.8373

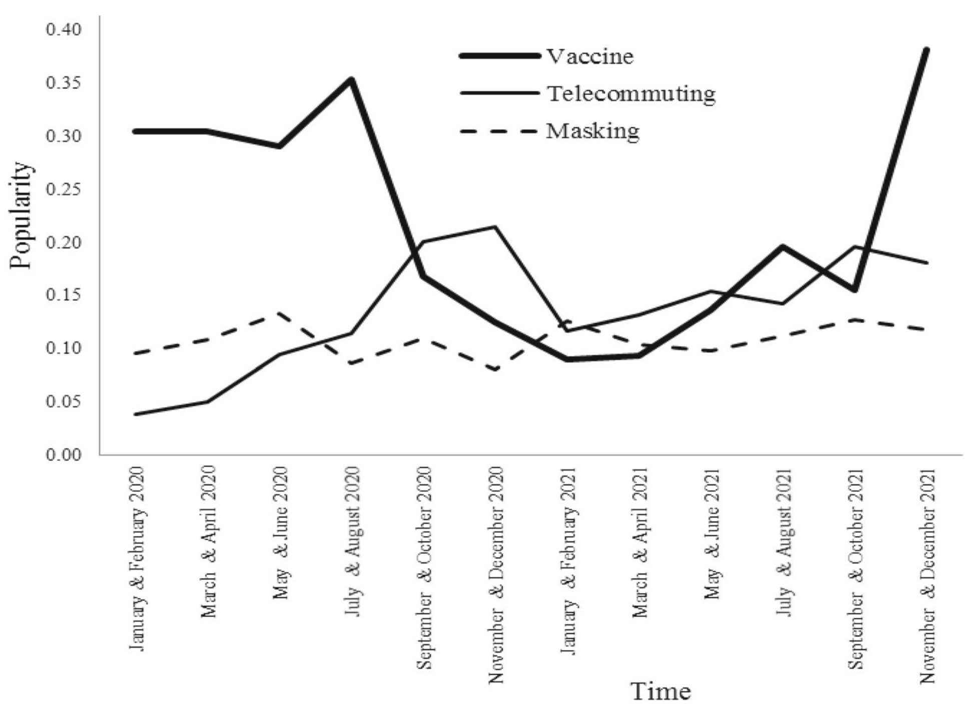


Fig. 5 The popularity of topics in the world

Table 4 The popularity of topics in USA

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3537	0.24	0.281	0.2938	0.2706	0.336	
Masking	0.1554	0.167	0.1261	0.0864	0.0617	0.094	
Telecommuting	0.0758	0.077	0.0816	0.0462	0.2043	0.035	
Death	0.0903	0.107	0.1009	0.1607	0.0417	0.133	
Economy	0.0331	0.089	0.0893	0.126	0.0788	0.099	
Spread	0.083	0.076	0.0595	0.0519	0.0535	0.081	
Politics	0.034	0.058	0.0683	0.0558	0.0594	0.059	
Students	0.0819	0.069	0.069	0.0867	0.089	0.055	
Voluntarily	0.0384	0.064	0.0664	0.0445	0.0173	0.04	
Plasma	0.0545	0.052	0.0578	0.048	0.1237	0.069	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3048	0.1209	0.1586	0.1245	0.0729	0.0317	2.5884
Masking	0.0726	0.1607	0.0709	0.265	0.0594	0.0698	1.751
Telecommuting	0.161	0.1272	0.1587	0.1939	0.2693	0.3212	1.3889
Death	0.0608	0.148	0.1949	0.0402	0.0392	0.0689	1.1862
Economy	0.0714	0.0923	0.0992	0.0312	0.068	0.0179	1.0393
Spread	0.0359	0.0521	0.0567	0.0961	0.2864	0.1072	0.9517
Politics	0.0891	0.0642	0.07	0.0998	0.0573	0.237	0.8949
Students	0.054	0.0674	0.0542	0.0539	0.0229	0.0686	0.7715
Voluntarily	0.0782	0.1172	0.0646	0.0808	0.0682	0.0156	0.7327
Plasma	0.0721	0.05	0.0723	0.0146	0.0565	0.0621	0.6953

Table 5 The popularity of topics in India

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1594	0.0805	0.1755	0.1840	0.1838	0.2516	
Death	0.1382	0.1495	0.1411	0.1023	0.176	0.1361	
Spread	0.0790	0.0969	0.1996	0.1779	0.1167	0.1133	
Plasma	0.0907	0.1036	0.1029	0.0476	0.0231	0.0826	
Telecommuting	0.0878	0.1001	0.0647	0.0833	0.0735	0.0848	
voluntarily	0.0967	0.1171	0.0771	0.0693	0.0423	0.0427	
Students	0.1088	0.1021	0.0487	0.0516	0.0924	0.0752	
Politics	0.0895	0.0755	0.0682	0.12909	0.1494	0.1000	
Masking	0.0530	0.0693	0.0538	0.0533	0.0447	0.0321	
Economy	0.0965	0.1053	0.0680	0.10116	0.098	0.0811	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3094	0.4068	0.4551	0.4374	0.2687	0.1590	3.0716
Death	0.0963	0.0522	0.0970	0.0594	0.1420	0.1001	1.3906
Spread	0.0376	0.0484	0.0986	0.0352	0.0309	0.1241	1.1693
Plasma	0.1323	0.0670	0.0353	0.1477	0.0955	0.1449	1.1588
Telecommuting	0.0429	0.0611	0.0587	0.0542	0.0622	0.1082	1.0736
voluntarily	0.0549	0.0558	0.0621	0.0939	0.1159	0.0294	0.9801
Students	0.0548	0.0228	0.0813	0.0437	0.0212	0.0081	0.8819
Politics	0.1030	0.1439	0.0367	0.0697	0.1060	0.0978	0.8576
Masking	0.0421	0.0739	0.0293	0.0401	0.1355	0.0779	0.7111
Economy	0.1264	0.0676	0.0454	0.0187	0.0216	0.1499	0.7054

Table 6 The popularity of topics in China

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.0538	0.0828	0.0413	0.0405	0.0593	0.0917	
Politics	0.1582	0.1433	0.0819	0.0817	0.0925	0.1334	
voluntarily	0.0710	0.1285	0.1162	0.1285	0.0846	0.1405	
Masking	0.1593	0.0862	0.1283	0.1085	0.0922	0.1016	
Economy	0.1343	0.0995	0.1165	0.0622	0.1034	0.0864	
Spread	0.1470	0.1572	0.1184	0.1933	0.1516	0.0651	
Telecommuting	0.1000	0.1094	0.0786	0.1283	0.1414	0.0908	
Plasma	0.0929	0.0832	0.1205	0.1291	0.1119	0.0823	
Death	0.0635	0.0985	0.1292	0.0847	0.039	0.0821	
Students	0.0196	0.0109	0.0686	0.0430	0.1241	0.1256	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1143	0.1834	0.251	0.2530	0.1039	0.0859	1.3614
Politics	0.1273	0.1265	0.0580	0.0404	0.0887	0.0237	1.3109
voluntarily	0.0935	0.0923	0.1482	0.0404	0.0609	0.0540	1.2956
Masking	0.0341	0.0424	0.0919	0.1484	0.1899	0.1124	1.1718
Economy	0.114	0.0964	0.1018	0.0814	0.0413	0.1238	1.1610
Spread	0.0626	0.0432	0.0229	0.0252	0.0286	0.0578	1.1605
Telecommuting	0.1156	0.0650	0.0551	0.0732	0.1134	0.0780	1.1590
Plasma	0.0904	0.0955	0.0637	0.0886	0.0805	0.1216	1.1561
Death	0.1240	0.1333	0.0757	0.1046	0.1140	0.1228	1.1493
Students	0.1237	0.1213	0.1313	0.1443	0.1782	0.2197	1.0735

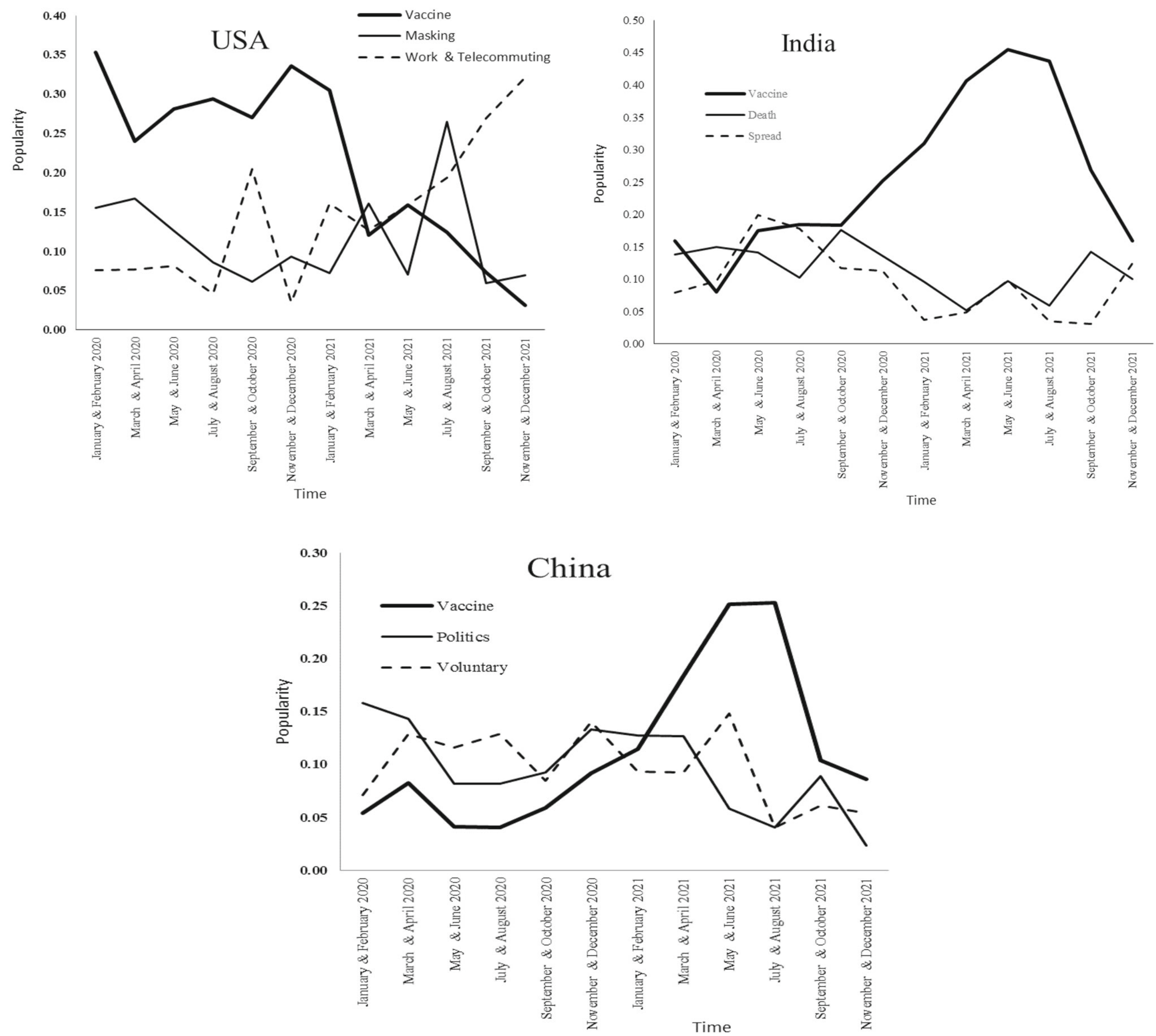


Fig. 6 Popularity trend of topics of 3 countries

As the proposed method focuses on the dynamic analysis of topics, a feature vector has been proposed to capture topic changes over time. Besides, selecting the optimal clustering and dimension reduction methods and clusters number resulted in the outperformance over previous methods.

5 Conclusion

Analyzing the time-dependent dynamics of topic modeling outputs using machine learning methods poses significant challenges. As such, the topic popularity over time metric

K-Means	3	4	5	6	7	8	9
Calinski-Harabasz Index	4.1457	3.3999	3.2775	2.9555	2.7696	2.7357	2.6998
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index	0.00035	0.00038	0.00034	0.00023	0.00021	0.0002	0.00019
Fuzzy C-Means	3	4	5	6	7	8	9
Calinski-Harabasz Index	4.8578	2.5538	2.8792	3.1532	2.9911	2.5595	2.3553
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index		0.00045		0.000309	0.00037	0.00023	0.0002
Hierarchical clustering	3	4	5	6	7	8	9
Calinski-Harabasz Index	4.0744	3.5842	3.4098	3.2491	3.1339	3.076	3.0354
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index	0.00044	0.00043	0.00041		0.00031	0.00036	0.00027

(a)

K-Means	3	4	5	6	7	8	9
Calinski-Harabasz Index				19.1283	17.5486	17.1941	16.0354
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index	0.00037	0.00039	0.00033	0.00028	0.00022	0.00024	0.00012
Fuzzy C-Means	3	4	5	6	7	8	9
Calinski-Harabasz Index		14.2934	10.0632	14.2934	14.2934	15.049	8.0979
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index				0.00041	0.00043	0.00034	0.0003
Hierarchical clustering	3	4	5	6	7	8	9
Calinski-Harabasz Index	16.536	11.1271	16.5916	14.7989	12.5156	11.5004	11.207
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index					0.00042	0.00043	0.0004

(b)

K-Means	3	4	5	6	7	8	9
Calinski-Harabasz Index							19.0647
Silhouette coefficient							
Davies-Bouldin Index							
Dunn Index			0.0004	0.00043	0.00038	0.00037	0.00022
Fuzzy C-Means	3	4	5	6	7	8	9
Calinski-Harabasz Index	2.4991	1.4377	1.7468	1.2995	1.2227	1.4377	1.3797
Silhouette coefficient	−0.2165	−0.1373	−1.2165	−0.2165	−0.2165	−0.2967	
Davies-Bouldin Index	2.728455	2.890971	4.219782	5.2779478	4.123216	5.277948	
Dunn Index	0.00034	0.00025	0.00033	0.00024	0.00027	0.00026	0.00017
Hierarchical clustering	3	4	5	6	7	8	9
Calinski-Harabasz Index	0.4849	1.4171	1.5901	1.8158	1.4832	1.2678	1.0729
Silhouette coefficient		−0.094	−0.1605	−0.1736	−0.215	−0.2391	−0.3465
Davies-Bouldin Index	7.766076	6.781978	6.977904	7.4292536	6.307828	4.65759	3.950288
Dunn Index	0.00041	0.000408	0.00035	0.00034	0.00023	0.00021	0.00012

(c)

Fig. 7 Visual selection of dimension reduction and clustering methods as well as number of clusters. **a** Without dimension reduction. **b** Dimension reduction with UMAP, **c** Dimension reduction with t-SNE

has been proposed, which is derived from the topic modeling outputs. Subsequently, this metric was used to model the feature vector for each entity (country) over time. Selecting the appropriate dimension reduction method and corresponding clustering algorithm was performed by identifying the best values of evaluation criteria, both visually and numerically. Finally, the proposed method has demonstrated superior performance over similar methods in terms of all evaluation criteria.

Based on the static analysis of the results, the Vaccine, Work and telecommuting, and Masking emerged as the first three topics of discussion across all countries during the years 2020 and 2021. While the topics discussed were similar across all countries, the proposed method revealed variations in the trends of these topics among different countries. As a result, the countries with similar top three topics were grouped into different clusters after dynamic clustering. Indeed, leveraging the dynamic analysis of topic modeling outputs in relation to various entities can contribute to a more comprehensive understanding of time-based topic changes. The proposed method can facilitate the analysis of entity dynamics across domains such as medicine, Politics, and psychology in natural language processing research, leading to improved research outcomes. This study solely examined the clustering of entities within the topic modeling output. Future research can explore the utilization of other machine learning methods, such as regression and classification. Finally, we proposed an indirect text clustering methodology that

Table 7 Normalized values of evaluation criteria, for K-means, FCM, and HC clustering methods, before dimension reduction and after dimension reduction with UMAP and t-SNE and for varying numbers of clusters

Clustering method	Evaluation criteria	Without dimension reduction							Dimension reduction with UMAP							Dimension reduction with t-SNE						
		Number of clusters (Sorted according to performance)																				
		4	3	5	6	7	8	9	4	3	5	6	7	8	9	4	3	6	7	5	8	9
K-Means	CHI	0.111	0.1394	0.1064	0.0941	0.087	0.0857	0.0843	0.8851	0.8179	0.7977	0.71	0.6498	0.6363	0.5922	0.8223	0.7901	0.8015	0.8232	0.8036	0.8379	0.7076
	SC	0.8328	0.8388	0.8323	0.8175	0.815	0.8128	0.8076	0.9904	1	0.9915	0.9525	0.962	0.9115	0.9189	0.9672	0.9831	0.9603	0.9544	0.9494	0.9395	0.8988
	DBI	0.8213	0.8043	0.8233	0.8699	0.8837	0.9007	0.9163	0.9607	0.9569	0.9578	0.9592	0.9591	0.9477	0.9693	0.9819	0.9789	0.9931	1	0.9791	0.9962	0.9857
	DI	0.5652	0.5	0.4783	0.2391	0.1957	0.1739	0.1522	0.587	0.5435	0.4565	0.3478	0.2174	0.2609	0	0.8696	0.8478	0.6739	0.5652	0.6087	0.5435	0.2174
	Mean	**0.5826**	0.5706	0.5601	0.5052	0.495	0.493	0.4901	**0.856**	0.8296	0.801	0.7424	0.6971	0.689	0.62	**0.9103**	0.9	0.8572	0.836	0.835	0.829	0.702
Clustering method	Evaluation criteria	Without dimension reduction							Dimension reduction with UMAP							Dimension reduction with t-SNE						
		Number of clusters (Sorted according to performance)																				
		3	5	4	6	7	8	9	3	4	5	7	6	8	9	3	9	4	7	6	8	5
FCM	CHI	0.1665	0.0912	0.0788	0.1016	0.0954	0.079	0.0712	1	0.5259	0.3648	0.5259	0.5259	0.5546	0.2899	0.0767	0.0341	0.0363	0.0281	0.031	0.0363	0.0481
	SC	0.8468	0.7999	0.8468	0.803	0.7519	0.7999	0.7999	0.9272	0.8348	0.8348	0.8887	0.8762	0.8348	0.8348	0.6407	0.8298	0.6914	0.6407	0.6407	0.5893	0
	DBI	0.7716	0.7664	0.8465	0.8662	0.7635	0.82	0.8465	0.9272	0.9766	0.9304	0.9766	0.98	0.9757	0.9766	0.7166	0.7427	0.6934	0.5182	0.3539	0.3539	0.5044
	DI	0.8283	0.913	0.7174	0.4109	0.5435	0.2391	0.1739	0.913	0.9348	1	0.6739	0.6304	0.4783	0.3913	0.4783	0.1087	0.2826	0.3261	0.2609	0.3043	0.4565
	Mean	**0.6533**	0.6426	0.6224	0.5454	0.539	0.485	0.4729	**0.9419**	0.818	0.783	0.7663	0.7531	0.711	0.623	**0.4781**	0.429	0.4259	0.378	0.322	0.321	0.3216
Clustering method	Evaluation criteria	Without dimension reduction							Dimension reduction with UMAP							Dimension reduction with t-SNE						
		Number of clusters (Sorted according to performance)																				
		6	3	4	5	8	7	9	3	5	6	4	7	8	9	4	3	5	8	6	7	9
HC	CHI	0.1053	0.1367	0.118	0.1114	0.0987	0.1009	0.0971	0.6113	0.6134	0.5451	0.4053	0.4582	0.4195	0.4083	0.0355	0	0.0421	0.0298	0.0507	0.038	0.0224
	SC	0.8217	0.833	0.8264	0.8259	0.8253	0.8243	0.8281	0.9561	0.9111	0.8885	0.8681	0.8555	0.8228	0.7976	0.7192	0.7523	0.6766	0.6262	0.6682	0.6417	0.5574
	DBI	0.8331	0.7947	0.8245	0.816	0.8857	0.8785	0.9091	0.9423	0.9148	0.9167	0.889	0.9659	0.9561	0.9627	0.14	0	0.1121	0.4422	0.0479	0.2074	0.5428
	DI	0.7826	0.6957	0.6739	0.6304	0.5217	0.413	0.3261	1	0.8478	0.8696	0.7826	0.6522	0.6739	0.6087	0.6261	0.6304	0.5	0.1957	0.4783	0.2391	0
	Mean	**0.6357**	0.615	0.6107	0.5959	0.583	0.554	0.5401	**0.8774**	0.822	0.805	0.736	0.733	0.718	0.694	**0.3802**	0.3457	0.333	0.323	0.3113	0.282	0.281

Table 8 Comparison of the proposed method with similar works

Method	Evaluation metric			
	CHI	SC	DBI	DI
The proposed method	22.0782	0.2931	0.7623	0.0003
Palanichamy et al. [4]	15.9772	0.1848	0.9517	0.0001
Sun and Yin [2] and Xie et al. [5]	20.1023	0.2149	0.8347	0.0003

converts text clustering to traditional machine learning clustering. In this regard, a fundamental question arises: Is it better to cluster the texts directly as in the previous research or extract the parameters using heuristic methods and then cluster based on them (like the approach we used in this paper)? Investigating this problem is an interesting fundamental future research.

Appendix 1

See Tables 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 and 37

Table 9 Popularity of topics in Australia

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.30137	0.327181	0.3085881	0.2652744	0.16662	0.28653
Telecommuting	0.1309654	0.11201	0.1346434	0.1099756	0.15836	0.13182
Death	0.0649379	0.059166	0.0494905	0.0773162	0.04372	0.06579
Politics	0.1048159	0.090771	0.0803008	0.1139747	0.08824	0.09178
Masking	0.0544781	0.067509	0.0982533	0.102866	0.09891	0.07104
Economy	0.0758335	0.067762	0.0749636	0.0675405	0.07785	0.05721
Spread	0.10438	0.113527	0.1055313	0.1021995	0.12983	0.09511
Plasma	0.0684245	0.070544	0.0412421	0.0428794	0.07731	0.05435
Students	0.054696	0.039444	0.0465793	0.0637636	0.06985	0.05769
voluntarily	0.0400959	0.052086	0.0604076	0.0542102	0.08931	0.08868

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.34456	0.295056	0.311067	0.33012	0.094595	0.21479	3.245764408
Telecommuting	0.09287	0.111643	0.115653	0.1393	0.179899	0.10563	1.52276709
Death	0.10193	0.118022	0.103689	0.17441	0.284628	0.28257	1.425665237
Politics	0.05125	0.060606	0.076271	0.05436	0.040541	0.02993	1.156509995
Masking	0.06738	0.075359	0.059821	0.04134	0.057432	0.03433	0.88283485
Economy	0.0722	0.070574	0.075274	0.05436	0.096284	0.07835	0.868191256
Spread	0.08862	0.088517	0.093719	0.06512	0.081926	0.08803	0.828713873
Plasma	0.07814	0.072169	0.046859	0.03681	0.043074	0.05106	0.735761103
Students	0.07305	0.064593	0.034895	0.04247	0.061655	0.04225	0.682860899
voluntarily	0.03001	0.043461	0.082752	0.06172	0.059966	0.07306	0.650931288

Table 10 Popularity of topics in UK

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.266530334	0.2194422	0.314607	0.1581532	0.1486523	0.2947761	
Telecommuting	0.030902068	0.1199754	0.096334	0.1863949	0.1815954	0.0961538	
Politics	0.115201091	0.1193601	0.110886	0.141945	0.096379	0.1228473	
Death	0.113610543	0.1054143	0.113649	0.1078094	0.1061802	0.097876	
Masking	0.129970461	0.1216161	0.067047	0.0739194	0.0876668	0.075775	
Spread	0.070892979	0.0754717	0.058759	0.0795678	0.1132589	0.0634328	
Plasma	0.094978414	0.0877769	0.064284	0.05722	0.0664307	0.0611366	
Economy	0.048852534	0.0486054	0.039234	0.0717092	0.0702423	0.0611366	
voluntarily	0.067711884	0.0787531	0.082151	0.0687623	0.0620746	0.0677382	
Students	0.061349693	0.0235849	0.053048	0.0545187	0.0675197	0.0591274	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.323829	0.111154	0.183792	0.1974742	0.091531	0.1120527	2.421993573
Telecommuting	0.168816	0.206098	0.115292	0.064868	0.065013	0.13371	1.465153255
Politics	0.121747	0.143188	0.079112	0.1096441	0.11805	0.1045198	1.403445717
Death	0.107264	0.108452	0.138447	0.1394948	0.171086	0.094162	1.382879312
Masking	0.041638	0.093786	0.075736	0.0820896	0.107784	0.1026365	1.059665787
Spread	0.050464	0.060594	0.112397	0.1458094	0.100941	0.0781544	1.009743306
Plasma	0.048427	0.069085	0.087313	0.0568312	0.092387	0.0630885	0.909405781
Economy	0.041865	0.075261	0.085384	0.1021814	0.127459	0.1374765	0.848959063
voluntarily	0.025798	0.068313	0.038591	0.0350172	0.058169	0.0612053	0.784468359
Students	0.070152	0.064068	0.083936	0.0665901	0.067579	0.1129944	0.714285846

Table 11 Popularity of topics in Canada

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3551	0.2760	0.2814	0.3208	0.3549	0.3640	
Telecommuting	0.1187	0.1377	0.1107	0.0826	0.0696	0.0754	
Masking	0.1150	0.1063	0.0949	0.0770	0.0949	0.0679	
Politics	0.1098	0.1215	0.1014	0.135	0.1004	0.0380	
Death	0.0536	0.0837	0.0922	0.0795	0.0905	0.0819	
Spread	0.0751	0.0540	0.0760	0.0687	0.0734	0.0958	
Students	0.0441	0.0842	0.1068	0.09270	0.0485	0.0598	
Plasma	0.0296	0.0456	0.0611	0.0565	0.0659	0.0625	
Economy	0.0278	0.0319	0.0357	0.0409	0.0485	0.0577	
voluntarily	0.0705	0.0587	0.0393	0.0458	0.0529	0.0965	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2467	0.2555	0.2455	0.2885	0.2570	0.3288	3.5747
Telecommuting	0.1376	0.1674	0.1389	0.1263	0.1756	0.1723	1.5135
Masking	0.1019	0.0789	0.1030	0.0918	0.0850	0.0970	1.1143
Politics	0.0616	0.0688	0.0636	0.0604	0.0786	0.0292	1.05064
Death	0.1145	0.0938	0.10943	0.133	0.0722	0.04518	0.9690
Spread	0.06374	0.0626	0.0764	0.07583	0.0924	0.0811	0.8956
Students	0.0583	0.0556	0.0445	0.04315	0.0402	0.04267	0.7898
Plasma	0.0826	0.0640	0.07933	0.06411	0.0860	0.0920	0.7629
Economy	0.0843	0.0868	0.1053	0.0838	0.0786	0.0811	0.7209
voluntarily	0.0482	0.0661	0.0335	0.0320	0.0338	0.0301	0.6081

Table 12 Popularity of topics in Pakistan

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2207	0.2227	0.2716	0.2866	0.3269	0.3570	
Death	0.1561	0.1350	0.1470	0.163	0.1346	0.0803	
Politics	0.1246	0.0789	0.0640	0.0529	0.0751	0.0598	
Masking	0.1111	0.1175	0.1366	0.0921	0.0664	0.0635	
Telecommuting	0.1096	0.1087	0.0882	0.1109	0.0664	0.0542	
Economy	0.0780	0.0842	0.0951	0.0904	0.0751	0.0878	
Students	0.0945	0.1070	0.0692	0.0631	0.0769	0.0934	
Spread	0.0420	0.0333	0.0380	0.0307	0.0437	0.0654	
Plasma	0.0150	0.0456	0.0467	0.0614	0.0786	0.0710	
Voluntary	0.0480	0.0666	0.0432	0.0477	0.0559	0.0672	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3543	0.2409	0.1646	0.1574	0.2126	0.1759	2.9915
Death	0.0837	0.1074	0.1107	0.0709	0.1142	0.0601	1.3644
Politics	0.0616	0.0997	0.1497	0.1635	0.1269	0.0925	1.1497
Masking	0.0859	0.0818	0.08982	0.0679	0.06031	0.0925	1.0768
Telecommuting	0.0660	0.1202	0.0868	0.04321	0.0691	0.1157	1.0658
Economy	0.0638	0.0383	0.0658	0.1450	0.1556	0.0972	1.0489
Students	0.0748	0.0997	0.0508	0.07099	0.0380	0.0601	1.0400
Spread	0.0682	0.0920	0.1736	0.1512	0.1206	0.1898	0.8991
Plasma	0.0815	0.06138	0.0539	0.0648	0.0539	0.0555	0.6895
Voluntary	0.0594	0.05882	0.0538	0.0648	0.0476	0.06019	0.673

Table 13 Popularity of topics in Japan

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2866521	0.279773	0.3044397	0.200495	0.22525	0.36638	
Death	0.1728665	0.115312	0.1057082	0.1212871	0.17574	0.05819	
Economy	0.1006565	0.124764	0.0866808	0.1287129	0.09158	0.04957	
Telecommuting	0.0634573	0.068053	0.1035941	0.1163366	0.08168	0.09267	
Politics	0.0284464	0.083176	0.0951374	0.0940594	0.10891	0.10129	
Masking	0.1094092	0.10397	0.0909091	0.0866337	0.07178	0.11853	
Spread	0.0962801	0.071834	0.0591966	0.0767327	0.09406	0.05172	
Voluntary	0.0568928	0.05482	0.0570825	0.0717822	0.05446	0.04957	
Students	0.059081	0.05293	0.0528	0.0371287	0.04703	0.06034	
Plasma	0.0262582	0.045369	0.0443	0.0668317	0.0495	0.05172	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.21203	0.169492	0.177686	0.10959	0.15528	0.1562	2.643308203
Death	0.0696	0.169492	0.095041	0.1643	0.1801	0.2109	1.63870442
Economy	0.10443	0.118644	0.095041	0.1643	0.0993	0.1093	1.30621847
Telecommuting	0.12975	0.118644	0.136364	0.13242	0.1304	0.1328	1.27322013
Politics	0.10759	0.118644	0.119835	0.10502	0.0993	0.1015	1.163060942
Masking	0.08544	0.067797	0.070248	0.08219	0.086957	0.0625	1.036374244
Spread	0.08544	0.061017	0.11157	0.06393	0.093168	0.0937	0.958701448
Voluntary	0.04114	0.047458	0.086777	0.05936	0.037267	0.0390	0.670211267
Students	0.07911	0.057627	0.049587	0.06393	0.055901	0.0546	0.655666275
Plasma	0.08544	0.071186	0.057851	0.05479	0.062112	0.0390	0.6545346

Table 14 Popularity of topics in Germany

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.2564103	0.302439	0.1491713	0.1041667	0.33784	0.34049
Death	0.0589744	0.080488	0.1353591	0.1631944	0.11149	0.1319
Masking	0.1230769	0.158537	0.2127072	0.1319444	0.0473	0.08282
Telecommuting	0.1205128	0.082927	0.0966851	0.1006944	0.0777	0.04908
Spread	0.0923077	0.060976	0.0552486	0.1215278	0.08446	0.08282
Economy	0.0564103	0.05122	0.0883978	0.0902778	0.10811	0.08589
Politics	0.0794872	0.092683	0.0718232	0.1041667	0.07095	0.06135
Students	0.0820513	0.060976	0.0635359	0.0833333	0.05743	0.05828
Plasma	0.0615385	0.065854	0.058011	0.0486111	0.04054	0.03067
Voluntary	0.0692308	0.043902	0.0690608	0.0520833	0.06419	0.07669

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.43732	0.420886	0.133929	0.12632	0.107527	0.12651	2.84299698
Death	0.11953	0.110759	0.147321	0.15263	0.112903	0.10241	1.483448976
Masking	0.08455	0.117089	0.089286	0.07895	0.053763	0.07831	1.426962945
Telecommuting	0.06997	0.041139	0.196429	0.17895	0.204301	0.26506	1.258331008
Spread	0.06706	0.060127	0.071429	0.06842	0.086022	0.06627	0.91665941
Economy	0.06706	0.044304	0.075893	0.04737	0.086022	0.04819	0.88798289
Politics	0.04956	0.056962	0.071429	0.05263	0.069892	0.05422	0.849137761
Students	0.04665	0.056962	0.053571	0.11053	0.11828	0.09639	0.835149815
Plasma	0.01458	0.037975	0.107143	0.11053	0.112903	0.12048	0.808835938
Voluntary	0.04373	0.053797	0.053571	0.07368	0.048387	0.04217	0.690494278

Table 15 Popularity of topics in France

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3242574	0.268097	0.2988506	0.2638436	0.32384	0.37313	
Telecommuting	0.0717822	0.104558	0.1408046	0.1530945	0.11744	0.10697	
Death	0.1237624	0.147453	0.1235632	0.1074919	0.1032	0.11194	
Masking	0.0816832	0.061662	0.0689655	0.0684039	0.08185	0.08955	
Spread	0.0792079	0.067024	0.066092	0.0912052	0.07829	0.05224	
Politics	0.0767327	0.101877	0.0747126	0.0977199	0.07473	0.04975	
Economy	0.0668317	0.075067	0.0718391	0.0488599	0.06762	0.06219	
Students	0.0569307	0.034853	0.0402299	0.0684039	0.04626	0.06468	
Plasma	0.0544554	0.067024	0.0545977	0.0521173	0.06406	0.02985	
Voluntary	0.0643564	0.072386	0.0603448	0.0488599	0.0427	0.0597	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.21429	0.19305	0.186147	0.11483	0.120192	0.32787	3.008402698
Telecommuting	0.13312	0.135135	0.142857	0.13876	0.100962	0.06967	1.415140587
Death	0.08766	0.07722	0.073593	0.08612	0.067308	0.03279	1.195240742
Masking	0.09416	0.142857	0.125541	0.15311	0.125	0.10246	1.142108148
Spread	0.07143	0.061776	0.103896	0.12919	0.139423	0.14754	1.087311237
Politics	0.08766	0.069498	0.069264	0.04785	0.0625	0.11885	0.931150028
Economy	0.08117	0.065637	0.051948	0.11483	0.139423	0.06967	0.915084128
Students	0.09416	0.104247	0.125541	0.10526	0.028846	0.02049	0.791747901
Plasma	0.06818	0.084942	0.069264	0.0622	0.115385	0.06967	0.789902185
Voluntary	0.06818	0.065637	0.051948	0.04785	0.100962	0.04098	0.723912347

Table 16 Popularity of topics in Irland

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2007722	0.274306	0.260223	0.2142857	0.20524	0.31799	
Telecommuting	0.1003861	0.128472	0.1078067	0.1269841	0.11354	0.1046	
Death	0.0772201	0.045139	0.0966543	0.1190476	0.0917	0.08368	
Masking	0.1158301	0.072917	0.0743494	0.1150794	0.13537	0.11715	
Politics	0.1196911	0.097222	0.063197	0.0714286	0.06987	0.08368	
Spread	0.1003861	0.128472	0.1152416	0.0833333	0.07424	0.05021	
Plasma	0.0772201	0.09375	0.0780669	0.0555556	0.0524	0.04184	
Economy	0.03861	0.041667	0.0557621	0.047619	0.08734	0.08787	
Voluntary	0.1119691	0.076389	0.0743494	0.0833333	0.0786	0.06276	
Students	0.0579151	0.041667	0.0743494	0.0833333	0.0917	0.05021	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.27347	0.191964	0.279835	0.20574	0.215962	0.1976	2.837396248
Telecommuting	0.13469	0.102679	0.082305	0.10048	0.107981	0.21557	1.425494297
Death	0.11837	0.138393	0.115226	0.08134	0.131455	0.09581	1.207514754
Masking	0.08163	0.058036	0.106996	0.09091	0.131455	0.10778	1.194035962
Politics	0.05306	0.116071	0.123457	0.10048	0.093897	0.05389	1.045946785
Spread	0.08163	0.09375	0.041152	0.11962	0.056338	0.08383	1.028200809
Plasma	0.02041	0.053571	0.082305	0.10048	0.098592	0.07186	0.840455949
Economy	0.08571	0.080357	0.049383	0.09569	0.098592	0.07186	0.826045722
Voluntary	0.06939	0.071429	0.069959	0.03828	0.00939	0.04192	0.807146045
Students	0.08163	0.09375	0.049383	0.06699	0.056338	0.05988	0.787763428

Table 17 Popularity of topics in Singapore

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.2322835	0.200772	0.0995025	0.205	0.24157	0.16514
Telecommuting	0.1023622	0.135135	0.1641791	0.115	0.14607	0.2156
Masking	0.1023622	0.104247	0.0995025	0.105	0.09551	0.10092
Death	0.0787402	0.081081	0.0497512	0.125	0.06742	0.06422
Spread	0.0944882	0.111969	0.1243781	0.07	0.06742	0.11009
Politics	0.0944882	0.104247	0.1044776	0.07	0.06742	0.09633
Economy	0.1023622	0.104247	0.1044776	0.075	0.06742	0.09174
Students	0.0551181	0.046332	0.0995025	0.105	0.11798	0.05505
Plasma	0.0787402	0.065637	0.1044776	0.07	0.06742	0.04587
Voluntary	0.0590551	0.046332	0.0497512	0.06	0.0618	0.05505

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2533	0.204545	0.29717	0.17838	0.370732	0.24103	2.689513459
Telecommuting	0.13122	0.145455	0.09434	0.11351	0.112195	0.18462	1.659680098
Masking	0.0905	0.095455	0.04717	0.13514	0.058537	0.07179	1.106123532
Death	0.0724	0.122727	0.09434	0.11351	0.082927	0.11282	1.064934337
Spread	0.09502	0.095455	0.09434	0.05405	0.02439	0.06154	1.003142436
Politics	0.09502	0.090909	0.113208	0.05405	0.029268	0.06154	0.980958982
Economy	0.09502	0.063636	0.056604	0.05405	0.102439	0.05128	0.968283662
Students	0.06335	0.05	0.056604	0.10811	0.058537	0.10256	0.91813703
Plasma	0.02262	0.054545	0.04717	0.05946	0.102439	0.06154	0.829307694
Voluntary	0.08145	0.077273	0.099057	0.12973	0.058537	0.05128	0.779918771

Table 18 Popularity of topics in UAE

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.2196262	0.240664	0.3318182	0.3175966	0.32995	0.33036
Telecommuting	0.0934579	0.087137	0.0909091	0.1072961	0.06091	0.0625
Death	0.1074766	0.116183	0.1136364	0.0600858	0.06599	0.10714
Spread	0.1074766	0.074689	0.0681818	0.055794	0.07614	0.08036
Economy	0.0654206	0.107884	0.0954545	0.1158798	0.08629	0.08482
Masking	0.0981308	0.087137	0.0545455	0.0686695	0.06091	0.08929
Politics	0.1028037	0.062241	0.0590909	0.0772532	0.06091	0.04911
Students	0.088785	0.091286	0.0772727	0.0472103	0.07107	0.04911
Voluntary	0.0514019	0.087137	0.0545455	0.0858369	0.08122	0.09375
Plasma	0.0654206	0.045643	0.0545455	0.0643777	0.1066	0.05357

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.36842	0.121795	0.120482	0.11486	0.133803	0.01626	2.645636895
Telecommuting	0.11404	0.173077	0.120482	0.17568	0.119718	0.17886	1.38406352
Death	0.09211	0.134615	0.108434	0.06757	0.056338	0.09756	1.127135068
Spread	0.08772	0.076923	0.114458	0.10135	0.098592	0.12195	1.063634843
Economy	0.0614	0.070513	0.084337	0.08784	0.06338	0.06504	1.037821473
Masking	0.0614	0.076923	0.108434	0.13514	0.091549	0.09756	1.0296879
Politics	0.05263	0.102564	0.084337	0.11486	0.133803	0.13821	0.988267045
Students	0.05263	0.096154	0.126506	0.08108	0.084507	0.10569	0.971298144
Voluntary	0.07895	0.076923	0.060241	0.03378	0.084507	0.0813	0.882862626
Plasma	0.0307	0.070513	0.072289	0.08784	0.133803	0.09756	0.869592485

Table 19 Popularity of topics in Mexico

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2076271	0.165854	0.0970464	0.1417004	0.12033	0.12105	
Death	0.1694915	0.112195	0.2742616	0.2510121	0.23237	0.21579	
Masking	0.1271186	0.165854	0.1012658	0.1214575	0.11618	0.12632	
Telecommuting	0.0805085	0.097561	0.0801688	0.097166	0.07884	0.08947	
Spread	0.0466102	0.102439	0.092827	0.0809717	0.06639	0.05263	
Politics	0.0889831	0.087805	0.092827	0.0607287	0.05394	0.05789	
Economy	0.0466102	0.04878	0.0886076	0.048583	0.08299	0.09474	
Voluntary	0.0805085	0.082927	0.0675105	0.0607287	0.07469	0.07368	
Students	0.0847458	0.092683	0.0548523	0.0769231	0.08714	0.10526	
Plasma	0.0677966	0.043902	0.0506329	0.0607287	0.08714	0.06316	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.20238	0.2	0.152542	0.16552	0.269504	0.29139	2.134947019
Death	0.15476	0.1	0.101695	0.06207	0.021277	0.03311	1.728029979
Masking	0.11905	0.147059	0.127119	0.09655	0.078014	0.03974	1.4365425
Telecommuting	0.11905	0.152941	0.161017	0.17241	0.141844	0.16556	1.365720072
Spread	0.08929	0.070588	0.084746	0.14483	0.12766	0.07947	1.03844655
Politics	0.05357	0.076471	0.084746	0.07586	0.12766	0.09934	0.959827495
Economy	0.07143	0.058824	0.042373	0.08276	0.113475	0.13907	0.91823727
Voluntary	0.07738	0.052941	0.084746	0.08276	0.028369	0.05298	0.828632577
Students	0.06548	0.082353	0.059322	0.02759	0.014184	0.01987	0.819223044
Plasma	0.04762	0.058824	0.101695	0.08966	0.078014	0.07947	0.770393494

Table 20 Popularity of topics in Italy

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.2863636	0.226601	0.2898551	0.3529412	0.28804	0.25654
Telecommuting	0.1681818	0.226601	0.2270531	0.1372549	0.13043	0.10995
Death	0.0863636	0.098522	0.0917874	0.1176471	0.11957	0.08901
Spread	0.0954545	0.103448	0.057971	0.0784314	0.06522	0.10471
Masking	0.0590909	0.054187	0.057971	0.0539216	0.06522	0.08377
Politics	0.0727273	0.044335	0.057971	0.0735294	0.06522	0.06806
Economy	0.0272727	0.049261	0.0338164	0.0539216	0.10326	0.09424
Students	0.0772727	0.064039	0.0869565	0.0490196	0.07065	0.08377
Voluntary	0.0681818	0.064039	0.0483092	0.0343137	0.05978	0.08901
Plasma	0.0590909	0.068966	0.0483092	0.0490196	0.03261	0.02094

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.29586	0.132353	0.160839	0.19595	0.130435	0.17431	2.790091597
Telecommuting	0.13609	0.095588	0.132867	0.08784	0.076087	0.08257	1.693554271
Death	0.11834	0.191176	0.132867	0.16892	0.25	0.22936	1.610516916
Spread	0.08284	0.088235	0.125874	0.13514	0.141304	0.16514	1.243761396
Masking	0.05325	0.088235	0.104895	0.09459	0.086957	0.09174	0.893836782
Politics	0.06509	0.088235	0.076923	0.08108	0.086957	0.09174	0.871870743
Economy	0.07692	0.102941	0.125874	0.06757	0.065217	0.0367	0.836994098
Students	0.04142	0.080882	0.062937	0.05405	0.021739	0.02752	0.720265728
Voluntary	0.09467	0.058824	0.027972	0.04054	0.043478	0.02752	0.682464643
Plasma	0.0355	0.073529	0.048951	0.07432	0.097826	0.07339	0.656643826

Table 21 Popularity of topics in Sweden

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.1404255	0.108491	0.1176471	0.125	0.13218	0.18848
Death	0.1531915	0.165094	0.1617647	0.1369048	0.14943	0.14136
Economy	0.2340426	0.245283	0.2254902	0.1845238	0.17241	0.12042
Students	0.0851064	0.084906	0.1029412	0.0892857	0.07471	0.05236
Masking	0.0553191	0.066038	0.0490196	0.0654762	0.03448	0.03665
Plasma	0.0638298	0.061321	0.0490196	0.0416667	0.06322	0.06283
Voluntary	0.0638298	0.061321	0.0539216	0.0714286	0.05172	0.03665
Politics	0.106383	0.117925	0.0980392	0.0714286	0.11494	0.13613
Telecommuting	0.0553191	0.04717	0.0441176	0.0595238	0.06322	0.04188
Spread	0.0425532	0.042453	0.0980392	0.1547619	0.14368	0.18325

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.15183	0.210227	0.246835	0.19876	0.25	0.17483	2.044706856
Death	0.15183	0.181818	0.126582	0.13043	0.159722	0.18182	1.839949948
Economy	0.13613	0.096591	0.075949	0.0559	0.027778	0.03497	1.666957166
Students	0.05759	0.102273	0.094937	0.06832	0.0625	0.09091	1.639682297
Masking	0.04712	0.056818	0.044304	0.06832	0.0625	0.05594	1.609481583
Plasma	0.05759	0.045455	0.025316	0.03727	0.027778	0.02098	0.96584073
Voluntary	0.05759	0.017045	0.025316	0.02484	0.041667	0.05594	0.641994092
Politics	0.16754	0.0625	0.082278	0.18012	0.256944	0.27273	0.561283011
Telecommuting	0.02094	0.017045	0.031646	0.03727	0.034722	0.02098	0.556268936
Spread	0.15183	0.210227	0.246835	0.19876	0.076389	0.09091	0.473835381

Table 22 Popularity of topics in Brazil

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Death	0.1774194	0.208589	0.2323944	0.1489362	0.13218	0.23841
Vaccine	0.2204301	0.159509	0.1197183	0.0851064	0.24713	0.15232
Telecommuting	0.0913978	0.09816	0.1267606	0.0851064	0.06897	0.08609
Masking	0.1075269	0.116564	0.0915493	0.1347518	0.12069	0.13245
Spread	0.0698925	0.08589	0.0704225	0.070922	0.09195	0.09272
Voluntary	0.1021505	0.110429	0.0915493	0.0992908	0.10345	0.07285
Politics	0.0537634	0.04908	0.0774648	0.106383	0.07471	0.06623
Economy	0.0806452	0.079755	0.0774648	0.0851064	0.05172	0.04636
Students	0.0430108	0.02454	0.0422535	0.106383	0.07471	0.06623
Plasma	0.0537634	0.067485	0.0704225	0.0780142	0.03448	0.04636

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Death	0.15591	0.188776	0.207447	0.25806	0.367347	0.2907	2.606178779
Vaccine	0.34946	0.316327	0.297872	0.07258	0.030612	0.05814	2.109201982
Telecommuting	0.03763	0.056122	0.101064	0.14516	0.122449	0.15116	1.170076285
Masking	0.05914	0.071429	0.06383	0.07258	0.020408	0.03488	1.025803027
Spread	0.08602	0.05102	0.037234	0.08871	0.091837	0.09302	0.990101729
Voluntary	0.05376	0.05102	0.037234	0.04839	0.061224	0.04651	0.929641443
Politics	0.05914	0.091837	0.079787	0.08871	0.091837	0.15116	0.877857125
Economy	0.05914	0.066327	0.074468	0.04839	0.081633	0.04651	0.797518466
Students	0.0914	0.056122	0.06383	0.08871	0.040816	0.03488	0.760736413
Plasma	0.04839	0.05102	0.037234	0.08871	0.091837	0.09302	0.73288475

Table 23 Popularity of topics in Iran

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Death	0.1933333	0.142857	0.2014925	0.1587302	0.19259	0.1666
Spread	0.1333333	0.135714	0.0970149	0.1507937	0.15556	0.1754
Vaccine	0.1133333	0.114286	0.1343284	0.1507937	0.13333	0.1052
Politics	0.0666667	0.078571	0.0746269	0.0714286	0.08148	0.0877
Telecommuting	0.08	0.092857	0.0746269	0.0873016	0.06667	0.0964
Economy	0.0733333	0.092857	0.0895522	0.0714286	0.05926	0.0964
Students	0.08	0.092857	0.0820896	0.0634921	0.07407	0.0438
Masking	0.1	0.071429	0.0820896	0.0952381	0.08148	0.0877
Voluntary	0.1	0.1	0.0895522	0.0873016	0.07407	0.0614
Plasma	0.06	0.078571	0.0746269	0.0634921	0.08148	0.0789

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Death	0.21368	0.2	0.247525	0.21348	0.289157	0.27536	2.494874489
Spread	0.17949	0.14	0.118812	0.17978	0.024096	0.04348	1.536604201
Vaccine	0.11111	0.11	0.118812	0.11236	0.144578	0.18841	1.533499335
Politics	0.07692	0.1	0.108911	0.07865	0.108434	0.11594	1.049355729
Telecommuting	0.11966	0.09	0.049505	0.07865	0.108434	0.10145	1.045641256
Economy	0.06838	0.12	0.128713	0.05618	0.072289	0.07246	1.000943413
Students	0.07692	0.06	0.069307	0.06742	0.096386	0.07246	0.87886753
Masking	0.05983	0.08	0.039604	0.06742	0.048193	0.04348	0.856476781
Voluntary	0.04274	0.06	0.029703	0.06742	0.036145	0.05797	0.806300745
Plasma	0.05128	0.04	0.089109	0.07865	0.072289	0.02899	0.797436519

Table 24 Popularity of topics in Russia

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2132353	0.284672	0.3043478	0.3112583	0.27215	0.29054	
Plasma	0.125	0.116788	0.1118012	0.0794702	0.07595	0.08784	
Telecommuting	0.0808824	0.058394	0.0745342	0.0993377	0.08228	0.07432	
Voluntary	0.1102941	0.094891	0.0869565	0.0794702	0.06962	0.08784	
Death	0.0588235	0.029197	0.0372671	0.0463576	0.08228	0.06757	
Politics	0.0955882	0.109489	0.0869565	0.0662252	0.07595	0.03378	
Spread	0.0882353	0.058394	0.068323	0.0596026	0.06329	0.03378	
Masking	0.0882353	0.080292	0.0621118	0.0529801	0.06962	0.06081	
Economy	0.0661765	0.072993	0.0434783	0.0331126	0.05063	0.02703	
Students	0.0735294	0.094891	0.1242236	0.1721854	0.15823	0.23649	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2679	0.241379	0.253846	0.3	0.278481	0.22973	3.247615433
Plasma	0.0457	0.075862	0.069231	0.06154	0.101266	0.09459	2.036061656
Telecommuting	0.0588	0.089655	0.076923	0.08462	0.101266	0.2027	1.083736918
Voluntary	0.0719	0.062069	0.053846	0.04615	0.101266	0.12162	1.045090318
Death	0.1111	0.075862	0.092308	0.08462	0.050633	0.04054	0.985921275
Politics	0.0588	0.055172	0.038462	0.04615	0.037975	0.05405	0.776561064
Spread	0.0719	0.068966	0.069231	0.05385	0.075949	0.02703	0.75863219
Masking	0.0653	0.027586	0.069231	0.05385	0.075949	0.02703	0.738544267
Economy	0.0588	0.048276	0.030769	0.04615	0.063291	0.05405	0.733049264
Students	0.1895	0.255172	0.246154	0.22308	0.113924	0.14865	0.594787615

Table 25 Popularity of topics in South Korea

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1328125	0.134454	0.1512605	0.1461538	0.14634	0.18182	
Masking	0.125	0.117647	0.1092437	0.1230769	0.11382	0.09091	
Telecommuting	0.09375	0.12605	0.1092437	0.1384615	0.09756	0.1	
Politics	0.0859375	0.067227	0.1008403	0.1	0.0813	0.1	
Death	0.140625	0.117647	0.1092437	0.0769231	0.13821	0.1	
Spread	0.1171875	0.109244	0.1176471	0.1	0.08943	0.10909	
Economy	0.0625	0.07563	0.0840336	0.0923077	0.10569	0.09091	
Plasma	0.0703125	0.067227	0.092437	0.1	0.09756	0.09091	
Students	0.1171875	0.117647	0.1008403	0.0846154	0.04878	0.07273	
Voluntary	0.0546875	0.067227	0.0252101	0.0384615	0.0813	0.06364	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.18182	0.12381	0.122449	0.14894	0.183099	0.15714	1.810094581
Masking	0.12397	0.095238	0.071429	0.11702	0.140845	0.11429	1.350502848
Telecommuting	0.09917	0.152381	0.132653	0.07447	0.126761	0.1	1.342483579
Politics	0.07438	0.133333	0.132653	0.11702	0.112676	0.21429	1.319655147
Death	0.09917	0.104762	0.091837	0.07447	0.056338	0.04286	1.181586877
Spread	0.09091	0.07619	0.091837	0.09574	0.098592	0.08571	1.152085665
Economy	0.10744	0.104762	0.081633	0.11702	0.056338	0.1	1.078263585
Plasma	0.10744	0.104762	0.081633	0.05319	0.084507	0.07143	1.021406109
Students	0.04959	0.047619	0.071429	0.08511	0.042254	0.05714	0.894935197
Voluntary	0.06612	0.057143	0.122449	0.11702	0.098592	0.05714	0.848986412

Table 26 Popularity of topics in Spain

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1678832	0.197183	0.2	0.1261261	0.12264	0.19835	
Death	0.1240876	0.133803	0.144	0.1801802	0.15094	0.1405	
Telecommuting	0.080292	0.084507	0.104	0.0900901	0.10377	0.06612	
Masking	0.1094891	0.091549	0.088	0.1081081	0.10377	0.1157	
Politics	0.0875912	0.070423	0.088	0.1171171	0.11321	0.08264	
Economy	0.1021898	0.091549	0.08	0.0720721	0.06604	0.04959	
Voluntary	0.1021898	0.084507	0.064	0.0990991	0.0566	0.06612	
Students	0.0875912	0.098592	0.064	0.0540541	0.04717	0.09917	
Spread	0.1094891	0.098592	0.096	0.0990991	0.15094	0.08264	
Plasma	0.0291971	0.049296	0.072	0.0540541	0.08491	0.09917	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.20588	0.176471	0.223404	0.23529	0.246575	0.25	2.34980771
Death	0.11765	0.127451	0.12766	0.16471	0.178082	0.16176	1.750820246
Telecommuting	0.08824	0.137255	0.138298	0.12941	0.109589	0.19118	1.322743735
Masking	0.12745	0.078431	0.095745	0.04706	0.09589	0.08824	1.149434082
Politics	0.12745	0.107843	0.085106	0.05882	0.082192	0.07353	1.093928291
Economy	0.04902	0.068627	0.095745	0.08235	0.082192	0.04412	1.019501849
Voluntary	0.05882	0.04902	0.074468	0.09412	0.041096	0.05882	0.88348977
Students	0.07843	0.058824	0.074468	0.09412	0.041096	0.05882	0.856340263
Spread	0.04902	0.078431	0.031915	0.08235	0.082192	0.05882	0.848863688
Plasma	0.09804	0.117647	0.053191	0.01176	0.041096	0.01471	0.725070366

Table 27 Popularity of topics in Switzerland

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.1891892	0.23	0.172043	0.2183	0.2471	0.2432
Telecommuting	0.1171171	0.11	0.1290323	0.1264	0.1348	0.2162
Masking	0.0720721	0.12	0.1182796	0.1048	0.1343	0.0270
Death	0.1171171	0.11	0.1290323	0.1034	0.0785	0.0405
Economy	0.1081081	0.11	0.0967742	0.1371	0.0561	0.0818
Politics	0.0990991	0.09	0.1075269	0.0344	0.0568	0.0808
Plasma	0.0990991	0.09	0.1075269	0.0914	0.0785	0.0810
Spread	0.0630631	0.05	0.0645161	0.0348	0.0642	0.0540
Students	0.0720721	0.05	0.0430108	0.0808	0.0674	0.1081
Voluntary	0.0630631	0.04	0.0322581	0.0685	0.0786	0.0675

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.32051	0.354839	0.196078	0.37313	0.387097	0.3272	3.25899105
Telecommuting	0.11538	0.112903	0.098039	0.0597	0.129032	0.1818	1.530512823
Masking	0.07692	0.129032	0.078431	0.04478	0.096774	0.0909	1.092504517
Death	0.07692	0.048387	0.098039	0.07463	0.064516	0.0901	1.032191352
Economy	0.11538	0.096774	0.0524	0.08955	0.032258	0.0363	1.019230471
Politics	0.08974	0.096774	0.0588	0.07463	0.096774	0.0545	0.939657422
Plasma	0.05128	0.048387	0.07831	0.02985	0.048387	0.0363	0.850645296
Spread	0.03846	0.032258	0.1375	0.1194	0.080645	0.1090	0.84101477
Students	0.03846	0.032258	0.078431	0.07463	0.048387	0.0366	0.729595008
Voluntary	0.07692	0.048387	0.117647	0.0597	0.016129	0.0636	0.705657291

Table 28 Popularity of topics in Turkey

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1854839	0.176471	0.2173913	0.1627907	0.15476	0.16092	
Death	0.1209677	0.098039	0.1130435	0.1162791	0.10714	0.13793	
Telecommuting	0.1129032	0.107843	0.0869565	0.0930233	0.08333	0.06897	
Economy	0.1209677	0.137255	0.1043478	0.127907	0.19048	0.11494	
Spread	0.1290323	0.098039	0.1304348	0.1627907	0.14286	0.12644	
Politics	0.0887097	0.088235	0.0869565	0.0930233	0.08333	0.11494	
Masking	0.0887097	0.137255	0.0869565	0.0348837	0.05952	0.06897	
Students	0.0564516	0.04902	0.0782609	0.0697674	0.08333	0.04598	
Voluntary	0.0645161	0.04902	0.0347826	0.0348837	0.02381	0.06897	
Plasma	0.0322581	0.058824	0.0608696	0.1046512	0.07143	0.09195	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.25926	0.25	0.291667	0.29851	0.290323	0.3617	2.809276003
Death	0.09877	0.069444	0.097222	0.08955	0.112903	0.08511	1.346143243
Telecommuting	0.06173	0.097222	0.125	0.10448	0.096774	0.06383	1.28952785
Economy	0.11111	0.111111	0.041667	0.10448	0.096774	0.08511	1.246397344
Spread	0.11111	0.097222	0.111111	0.08955	0.048387	0.04255	1.102057201
Politics	0.07407	0.097222	0.055556	0.02985	0.048387	0.04255	0.922944748
Masking	0.08642	0.083333	0.041667	0.07463	0.096774	0.06383	0.902843498
Students	0.03704	0.027778	0.083333	0.1194	0.080645	0.12766	0.858665746
Voluntary	0.09877	0.083333	0.097222	0.07463	0.096774	0.06383	0.790528942
Plasma	0.06173	0.083333	0.055556	0.01493	0.032258	0.06383	0.731615425

Table 29 Popularity of topics in Netherlands

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.2169811	0.252525	0.2407407	0.3052632	0.29703	0.36364
Telecommuting	0.1415094	0.141414	0.1203704	0.0947368	0.11881	0.09091
Politics	0.0754717	0.111111	0.0833333	0.0631579	0.07921	0.07955
Masking	0.1037736	0.10101	0.1018519	0.0947368	0.07921	0.06818
Death	0.1320755	0.10101	0.1203704	0.0736842	0.08911	0.05682
Spread	0.0849057	0.111111	0.0925926	0.0842105	0.08911	0.07955
Economy	0.0754717	0.060606	0.0740741	0.0631579	0.06931	0.06818
Plasma	0.0849057	0.060606	0.0648148	0.0842105	0.05941	0.04545
Students	0.0377358	0.030303	0.0462963	0.0421053	0.0495	0.10227
Voluntary	0.0471698	0.030303	0.0555556	0.0947368	0.06931	0.04545

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.32584	0.327869	0.208333	0.2766	0.257143	0.33333	3.405293167
Telecommuting	0.10112	0.081967	0.125	0.10638	0.085714	0.12121	1.329151954
Politics	0.10112	0.098361	0.145833	0.17021	0.142857	0.12121	1.271427027
Masking	0.08989	0.114754	0.0625	0.08511	0.057143	0.09091	1.049062189
Death	0.06742	0.016393	0.041667	0.06383	0.085714	0.06061	0.923953233
Spread	0.08989	0.098361	0.0625	0.04255	0.028571	0.06061	0.908693219
Economy	0.04494	0.04918	0.083333	0.10638	0.114286	0.09091	0.899833742
Plasma	0.06742	0.065574	0.041667	0.02128	0.085714	0.06061	0.752407523
Students	0.07865	0.081967	0.125	0.04255	0.085714	0.0303	0.741650658
Voluntary	0.03371	0.065574	0.104167	0.08511	0.057143	0.0303	0.718527288

Table 30 Popularity of topics in Belgium

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1086957	0.141176	0.1770833	0.2261905	0.29167	0.37143	
Telecommuting	0.1304348	0.105882	0.1145833	0.0952381	0.125	0.08571	
Death	0.1195652	0.152941	0.125	0.0952381	0.08333	0.05714	
Masking	0.1521739	0.117647	0.1145833	0.1547619	0.08333	0.08571	
Economy	0.1086957	0.094118	0.09375	0.0833333	0.08333	0.05714	
Spread	0.0652174	0.082353	0.0833333	0.0595238	0.09722	0.08571	
Politics	0.0869565	0.070588	0.0833333	0.0833333	0.06944	0.05714	
Plasma	0.0652174	0.082353	0.0520833	0.0714286	0.05556	0.04286	
Students	0.076087	0.070588	0.0729167	0.0833333	0.08333	0.04286	
Voluntary	0.0869565	0.082353	0.0833333	0.047619	0.02778	0.11429	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.43478	0.243243	0.258065	0.23077	0.347826	0.16667	2.997593523
Telecommuting	0.10145	0.135135	0.193548	0.11538	0.043478	0.125	1.370848524
Death	0.10145	0.135135	0.129032	0.07692	0.130435	0.08333	1.289528541
Masking	0.04348	0.054054	0.032258	0.11538	0.043478	0.16667	1.163533751
Economy	0.04348	0.054054	0.064516	0.11538	0.086957	0.04167	1.026432996
Spread	0.10145	0.081081	0.032258	0.07692	0.043478	0.08333	0.926429071
Politics	0.02899	0.108108	0.064516	0.07692	0.130435	0.16667	0.891887075
Plasma	0.05797	0.108108	0.096774	0.07692	0.043478	0.08333	0.836082923
Students	0.05797	0.054054	0.096774	0.03846	0.086957	0.04167	0.804999657
Voluntary	0.02899	0.027027	0.032258	0.07692	0.043478	0.04167	0.692663939

Table 31 Popularity of topics in Denmark

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.2291667	0.22093	0.2191781	0.2714286	0.35714	0.43836	
Masking	0.1354167	0.116279	0.1506849	0.1857143	0.08571	0.06849	
Politics	0.1145833	0.151163	0.1917808	0.1428571	0.08571	0.05479	
Telecommuting	0.0833333	0.116279	0.109589	0.0857143	0.1	0.05479	
Spread	0.09375	0.081395	0.0684932	0.0571429	0.07143	0.06849	
Death	0.09375	0.093023	0.0684932	0.0428571	0.04286	0.05479	
Economy	0.0729167	0.069767	0.0684932	0.1	0.08571	0.0411	
Students	0.0729167	0.05814	0.0547945	0.0428571	0.05714	0.06849	
Plasma	0.0520833	0.05814	0.0410959	0.0571429	0.08571	0.0411	
Voluntary	0.0520833	0.034884	0.0273973	0.0142857	0.02857	0.10959	
Topic	**2021**						**P_K**
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.45161	0.433333	0.176471	0.17241	0.269231	0.3	3.539263962
Masking	0.04839	0.033333	0.058824	0.10345	0.153846	0.05	1.190140779
Politics	0.04839	0.016667	0.117647	0.06897	0.038462	0.15	1.181020773
Telecommuting	0.08065	0.1	0.176471	0.10345	0.153846	0.15	1.31412043
Spread	0.09677	0.066667	0.088235	0.06897	0.115385	0.05	0.926729366
Death	0.04839	0.066667	0.088235	0.13793	0.076923	0.05	0.863918382
Economy	0.01613	0.033333	0.088235	0.13793	0.076923	0.05	0.840539206
Students	0.08065	0.066667	0.088235	0.06897	0.038462	0.05	0.747318051
Plasma	0.06452	0.066667	0.088235	0.06897	0.038462	0.1	0.762116937
Voluntary	0.06452	0.116667	0.029412	0.06897	0.038462	0.05	0.634832115

Table 32 Popularity of topics in Chile

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.1298701	0.169014	0.2463768	0.3134328	0.33333	0.39474
Death	0.1428571	0.197183	0.173913	0.1492537	0.1	0.05263
Telecommuting	0.1038961	0.084507	0.115942	0.0895522	0.08333	0.07895
Spread	0.1038961	0.084507	0.115942	0.1044776	0.06667	0.02632
Politics	0.1168831	0.112676	0.0724638	0.0447761	0.06667	0.06579
Economy	0.0649351	0.070423	0.0869565	0.0895522	0.06667	0.07895
Masking	0.1168831	0.112676	0.0724638	0.0447761	0.06667	0.05263
Students	0.0909091	0.070423	0.057971	0.0746269	0.06667	0.03947
Plasma	0.0649351	0.056338	0.0289855	0.0447761	0.11667	0.10526
Voluntary	0.0649351	0.042254	0.0289855	0.0447761	0.03333	0.10526

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.35849	0.230769	0.235294	0.25806	0.26087	0.18182	3.112070215
Death	0.03774	0.025641	0.117647	0.09677	0.043478	0.13636	1.297786952
Telecommuting	0.07547	0.076923	0.117647	0.16129	0.173913	0.13636	1.27347862
Spread	0.03774	0.128205	0.117647	0.06452	0.130435	0.18182	1.162162373
Politics	0.0566	0.102564	0.088235	0.12903	0.086957	0.04545	0.988101697
Economy	0.07547	0.102564	0.088235	0.06452	0.043478	0.09091	0.922654971
Masking	0.0566	0.076923	0.088235	0.06452	0.086957	0.04545	0.884786647
Students	0.09434	0.102564	0.088235	0.06452	0.086957	0.04545	0.882136073
Plasma	0.07547	0.051282	0.029412	0.03226	0.043478	0.09091	0.739775475
Voluntary	0.13208	0.102564	0.029412	0.06452	0.043478	0.04545	0.737046978

Table 33 Popularity of topics in Saudi Arabia

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1375	0.184615	0.254902	0.2040816	0.2244	0.2857	
Telecommuting	0.175	0.138462	0.0980392	0.122449	0.1020	0.0952	
Spread	0.1125	0.107692	0.0980392	0.0816327	0.10204	0.1190	
Politics	0.1	0.138462	0.0588235	0.0816327	0.10204	0.0714	
Death	0.0875	0.076923	0.0980392	0.0816327	0.08163	0.0714	
Economy	0.1	0.107692	0.0980392	0.0612245	0.08163	0.0713	
Plasma	0.1	0.076923	0.0784314	0.0612245	0.08163	0.1190	
Masking	0.1	0.076923	0.0588235	0.0816327	0.08163	0.0476	
Voluntary	0.05	0.030769	0.0588235	0.1428571	0.08163	0.0714	
Students	0.0375	0.061538	0.0980392	0.0816327	0.06122	0.0476	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.3095	0.342105	0.3333	0.32	0.3333	0.375	3.304598799
Telecommuting	0.1190	0.105263	0.090909	0.08	0.125	0.04167	1.33164867
Spread	0.1428	0.105263	0.090909	0.08	0.125	0.16667	1.29311518
Politics	0.0924	0.078947	0.151515	0.08	0.083333	0.04167	1.083087724
Death	0.0952	0.078947	0.090909	0.08	0.083333	0.04167	0.967250725
Economy	0.0476	0.052632	0.030303	0.08	0.041667	0.04167	0.910912663
Plasma	0.0714	0.026316	0.090909	0.08	0.083333	0.04167	0.835262913
Masking	0.0462	0.026316	0.030303	0.08	0.041667	0.04167	0.813904228
Voluntary	0.0472	0.078947	0.060606	0.04	0.041667	0.04167	0.746016938
Students	0.0238	0.105263	0.030303	0.08	0.041667	0.16667	0.714202161

Table 34 Popularity of topics in Peru

Topic	2020					
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$
Vaccine	0.1864407	0.218182	0.2765957	0.2040816	0.17391	0.275
Death	0.1525424	0.145455	0.1276596	0.1428571	0.08696	0.125
Telecommuting	0.1186441	0.145455	0.1276596	0.1632653	0.17391	0.1
Masking	0.0847458	0.109091	0.0638298	0.1020408	0.08696	0.075
Politics	0.0677966	0.090909	0.1276596	0.0408163	0.06522	0.1
Students	0.0677966	0.054545	0.0425532	0.0204082	0.06522	0.05
Spread	0.0847458	0.072727	0.0638298	0.0816327	0.1087	0.05
Voluntary	0.1186441	0.109091	0.0851064	0.0612245	0.04348	0.075
Economy	0.0677966	0.036364	0.0212766	0.1020408	0.08696	0.075
Plasma	0.0508475	0.018182	0.0638298	0.0816327	0.1087	0.075

Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.42857	0.285714	0.34375	0.33333	0.230769	0.2	3.156351195
Death	0.05714	0.095238	0.03125	0.08333	0.192308	0.2	1.439742135
Telecommuting	0.08571	0.095238	0.0625	0.125	0.076923	0.05	1.324311995
Masking	0.02857	0.047619	0.125	0.04167	0.153846	0.15	1.068367094
Politics	0.05714	0.071429	0.03125	0.08333	0.115385	0.05	0.900938371
Students	0.11429	0.166667	0.09375	0.08333	0.038462	0.05	0.880449809
Spread	0.05714	0.119048	0.0625	0.04167	0.038462	0.1	0.847018064
Voluntary	0.02857	0.071429	0.03125	0.08333	0.038462	0.05	0.811717588
Economy	0.05714	0.02381	0.125	0.04167	0.038462	0.1	0.795588982
Plasma	0.08571	0.02381	0.09375	0.08333	0.076923	0.05	0.775514766

Table 35 Popularity of topics in Portugal

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1794872	0.227273	0.1860465	0.2682927	0.33333	0.35484	
Telecommuting	0.1282051	0.136364	0.1860465	0.0487805	0.11111	0.03226	
Death	0.1282051	0.113636	0.0697674	0.1219512	0.08333	0.03226	
Politics	0.0769231	0.068182	0.1395349	0.1219512	0.08333	0.12903	
Masking	0.1282051	0.090909	0.1162791	0.0731707	0.05556	0.03226	
Spread	0.1025641	0.113636	0.0930233	0.1219512	0.02778	0.03226	
Voluntary	0.0512821	0.090909	0.0697674	0.097561	0.05556	0.03226	
Plasma	0.025641	0.045455	0.0697674	0.0731707	0.11111	0.16129	
Economy	0.0769231	0.045455	0.0232558	0.0487805	0.05556	0.12903	
Students	0.1025641	0.068182	0.0465116	0.0243902	0.08333	0.06452	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.32258	0.342857	0.314286	0.2963	0.26087	0.2	3.286160508
Telecommuting	0.12903	0.142857	0.057143	0.14815	0.086957	0.15	1.356901868
Death	0.0645	0.085714	0.114286	0.07407	0.173913	0.15	1.211654798
Politics	0.0967	0.028571	0.057143	0.11111	0.086957	0.05	1.049512702
Masking	0.0967	0.028571	0.085714	0.03704	0.086957	0.15	0.981431107
Spread	0.0645	0.057143	0.085714	0.11111	0.043478	0.05	0.903173428
Voluntary	0.0642	0.085714	0.085714	0.03704	0.043478	0.1	0.821879348
Plasma	0.0322	0.028571	0.057143	0.03704	0.130435	0.05	0.814756838
Economy	0.0977	0.114286	0.057143	0.07407	0.043478	0.05	0.813793178
Students	0.0322	0.085714	0.085714	0.07407	0.043478	0.05	0.760736226

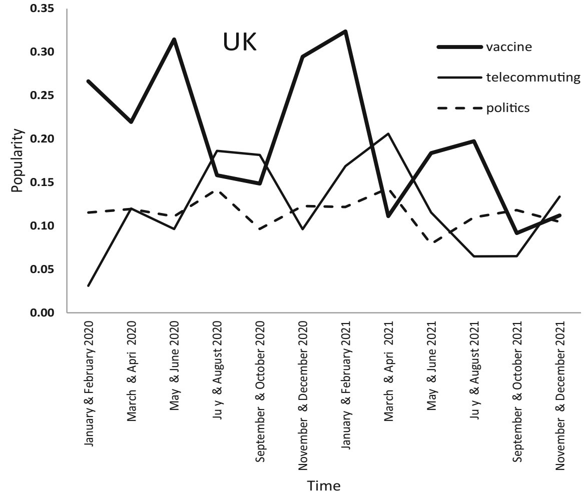
Table 36 Popularity of topics in Qatar

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.1944444	0.09375	0.16	0.225	0.27027	0.44444	
Telecommuting	0.1111111	0.15625	0.12	0.125	0.13514	0.11111	
Politics	0.1388889	0.125	0.12	0.1	0.08108	0.03704	
Death	0.1388889	0.125	0.12	0.1	0.08108	0.07407	
Spread	0.0833333	0.125	0.08	0.1	0.10811	0.07407	
Masking	0.1111111	0.15625	0.08	0.075	0.02703	0.03704	
Plasma	0.0555556	0.03125	0.08	0.025	0.05405	0.07407	
Students	0.0833333	0.0625	0.04	0.05	0.05405	0.03704	
Economy	0.0277778	0.0625	0.12	0.1	0.13514	0.07407	
Voluntary	0.0555556	0.0625	0.08	0.1	0.05405	0.03704	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.28571	0.275862	0.269231	0.28571	0.117647	0.22222	2.84429985
Telecommuting	0.14286	0.103448	0.115385	0.04762	0.117647	0.05556	1.341119053
Politics	0.03571	0.068966	0.038462	0.14286	0.117647	0.16667	1.207445498
Death	0.03571	0.034483	0.115385	0.09524	0.176471	0.11111	1.172319217
Spread	0.10714	0.068966	0.076923	0.04762	0.058824	0.05556	0.985545099
Masking	0.07143	0.034483	0.038462	0.09524	0.117647	0.11111	0.954794309
Plasma	0.14286	0.103448	0.153846	0.09524	0.058824	0.05556	0.929702436
Students	0.10714	0.137931	0.076923	0.09524	0.058824	0.05556	0.880182017
Economy	0.03571	0.068966	0.038462	0.04762	0.058824	0.11111	0.858538573
Voluntary	0.03571	0.103448	0.076923	0.04762	0.117647	0.05556	0.826053947

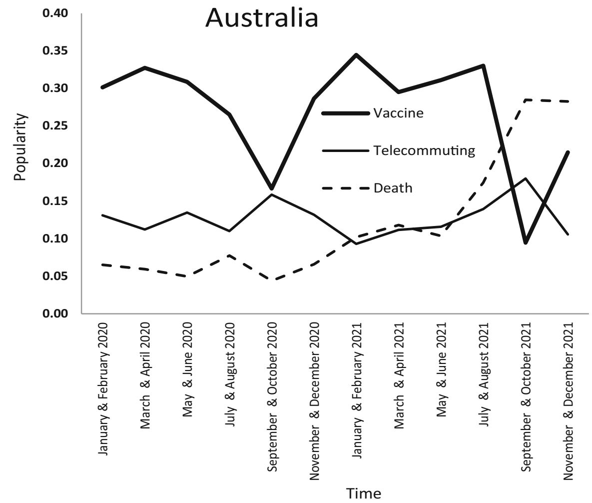
Table 37 Popularity of topics in Ecuador

Topic	2020						
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.0168	0.0126	0.0126	0.0210	0.0211	0.0253	
Telecommuting	0.0122	0.0168	0.0126	0.0168	0.0168	0.0084	
Death	0.0084	0.0084	0.0126	0.0126	0.0042	0.0084	
Students	0.0084	0.0084	0.0126	0.0084	0.0042	0.0084	
Masking	0.0084	0.0084	0.0084	0.0084	0.0042	0.0042	
Spread	0.0084	0.0126	0.0084	0.0084	0.0084	0.0042	
Plasma	0.0084	0.0042	0.0042	0.0084	0.0042	0.0084	
Politics	0.0042	0.0042	0.00843	0.0084	0.0084	0.0042	
Economy	0.0084	0.0042	0.0084	0.0042	0.0042	0.0084	
Voluntary	0.0042	0.0042	0.0042	0.0084	0.0084	0.0042	
Topic	2021						P_K
	$p_{\text{January \& February}}$	$p_{\text{March \& April}}$	$p_{\text{May \& June}}$	$p_{\text{July \& August}}$	$p_{\text{September \& October}}$	$p_{\text{November \& December}}$	
Vaccine	0.0337	0.0295	0.0084	0.0126	0.0042	0.0042	0.2025
Telecommuting	0.0126	0.0084	0.0084	0.0042	0.0042	0.0084	0.1308
Death	0.0042	0.0042	0.0126	0.0126	0.0042	0.0042	0.0970
Students	0.0084	0.0042	0.0042	0.0042	0.0084	0.0042	0.0843
Masking	0.0084	0.0126	0.0042	0.0042	0.0042	0.0042	0.0801
Spread	0.0042	0.0042	0.0042	0.0084	0.0042	0.0042	0.0801
Plasma	0.0084	0.0042	0.0042	0.0084	0.0084	0.0042	0.0759
Politics	0.0084	0.0042	0.0084	0.0042	0.0042	0.0084	0.0759
Economy	0.0084	0.0042	0.0084	0.0084	0.0042	0.0042	0.0759
Voluntary	0.0084	0.0042	0.0042	0.0042	0.0084	0.0084	0.0717

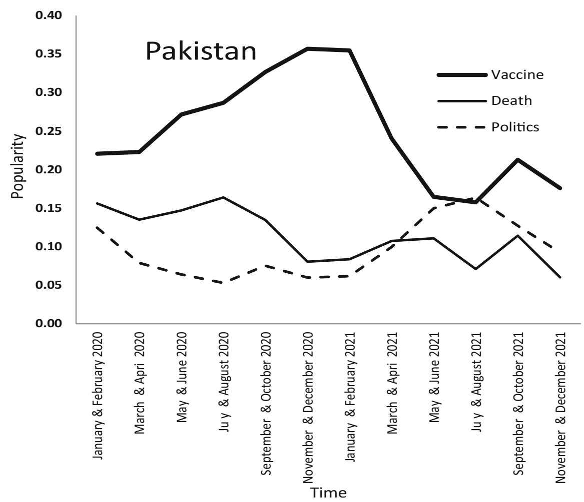
Appendix 2



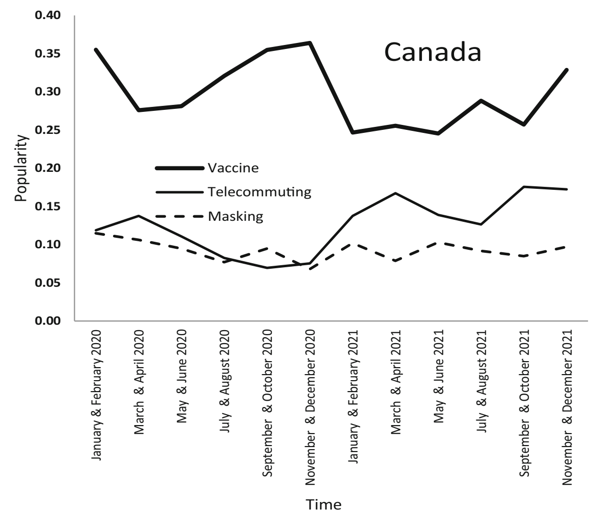
UK
vaccine
telecommuting
politics
Popularity
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time



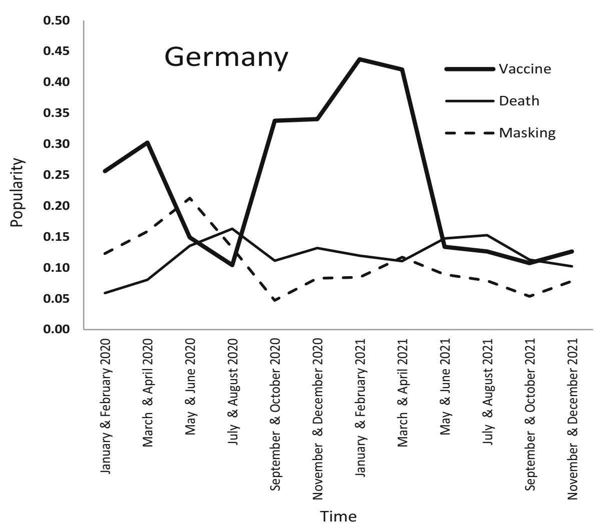
Australia
Vaccine
Telecommuting
Death
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time



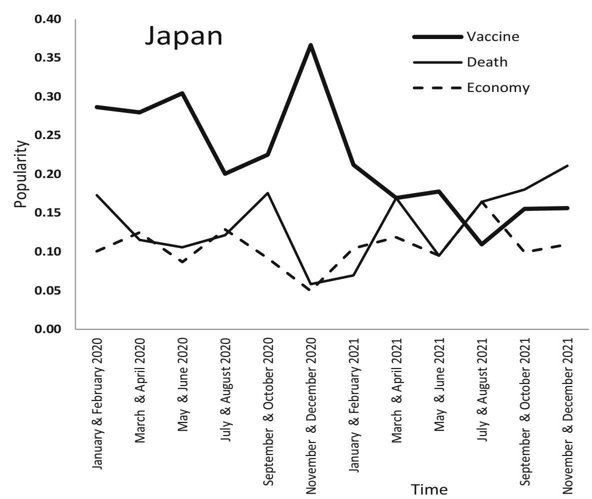
Pakistan
Vaccine
Death
Politics
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time



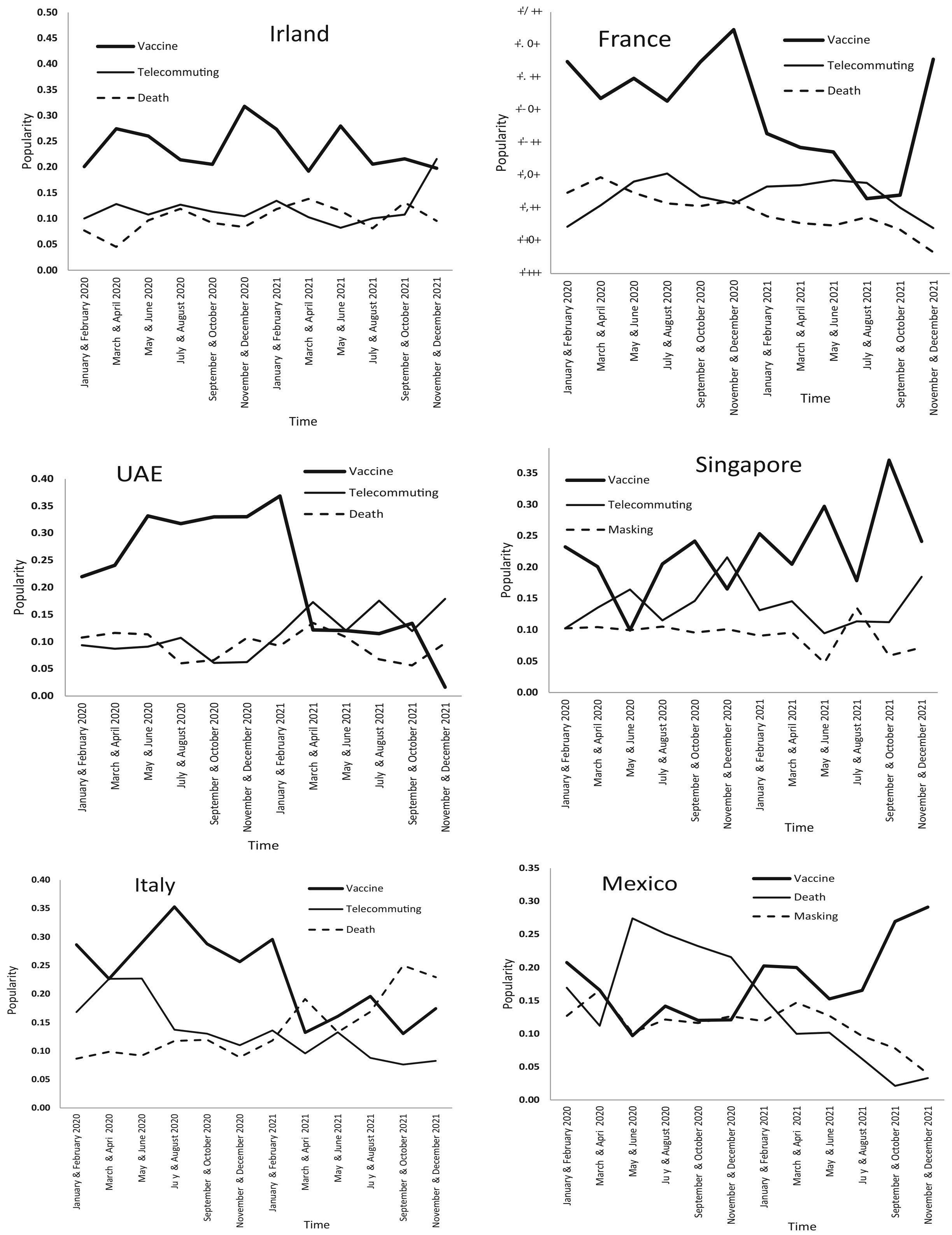
Canada
Vaccine
Telecommuting
Masking
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time

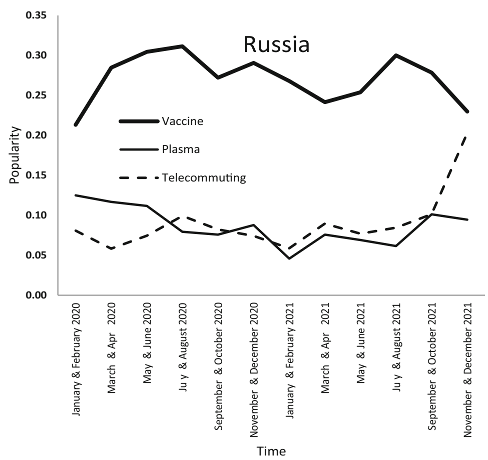


Germany
Vaccine
Death
Masking
Popularity
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time

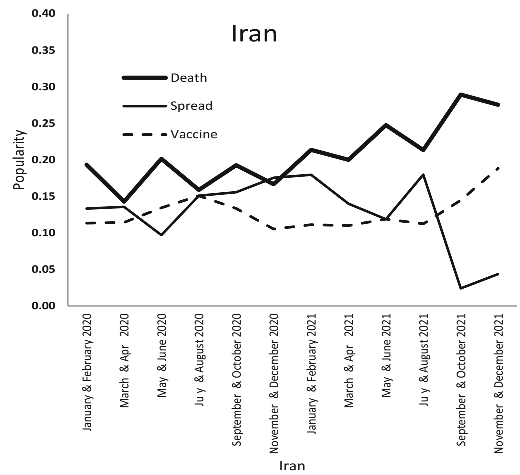


Japan
Vaccine
Death
Economy
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time

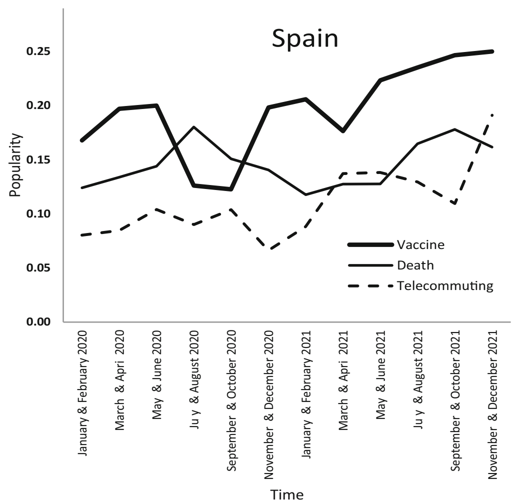

Irland
Vaccine
Telecommuting
Death
Popularity
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time
France
Vaccine
Telecommuting
Death
Popularity
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time
UAE
Vaccine
Telecommuting
Death
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time
Singapore
Vaccine
Telecommuting
Masking
Popularity
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time
Italy
Vaccine
Telecommuting
Death
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time
Mexico
Vaccine
Death
Masking
Popularity
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time



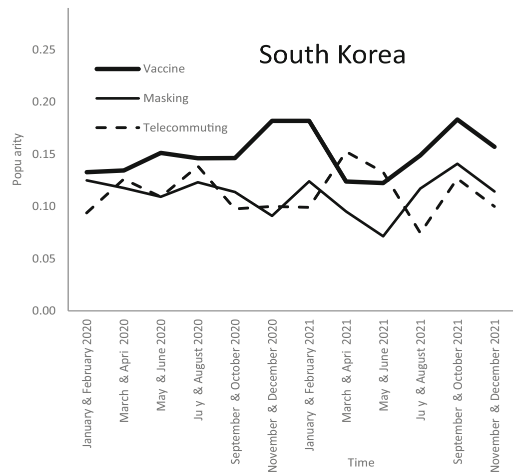
Russia
Vaccine
Plasma
Telecommuting
Popularity
Time



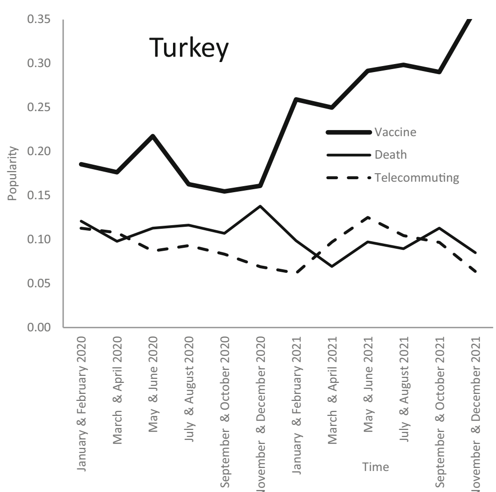
Iran
Death
Spread
Vaccine
Popularity
Iran



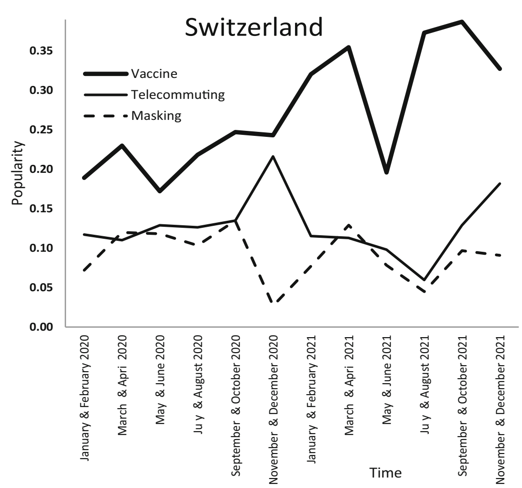
Spain
Vaccine
Death
Telecommuting
Popularity
Time



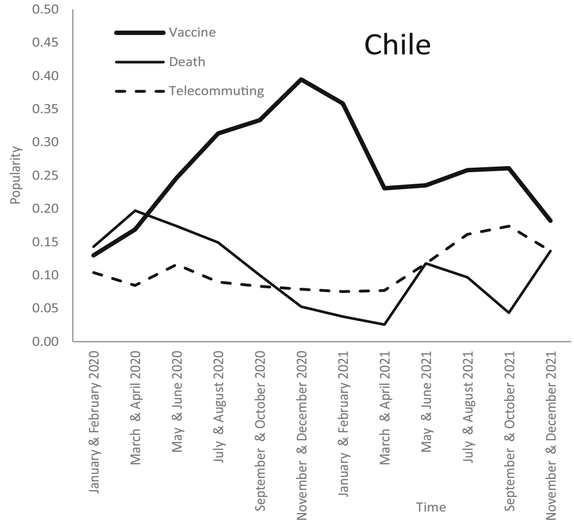
South Korea
Vaccine
Masking
Telecommuting
Popu arity
Time



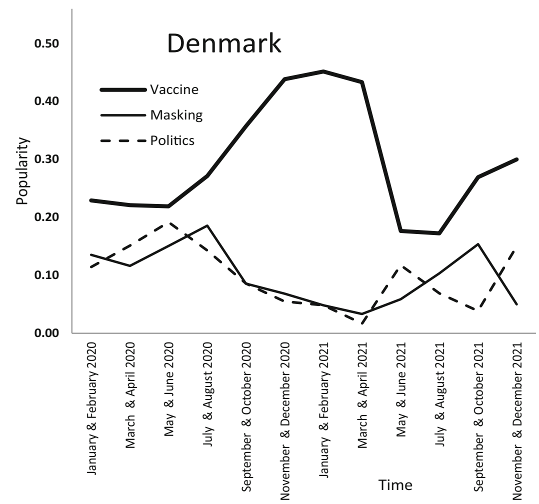
Turkey
Vaccine
Death
Telecommuting
Popularity
Time



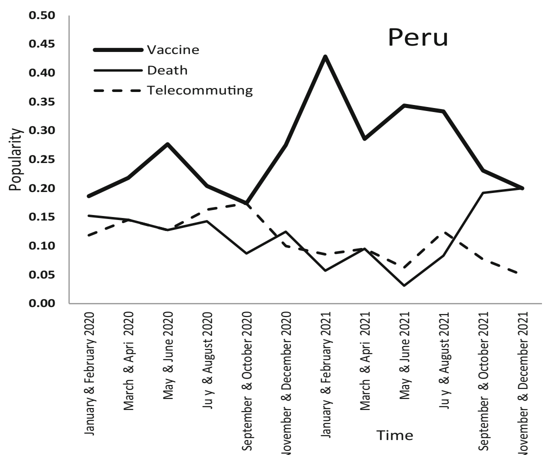
Switzerland
Vaccine
Telecommuting
Masking
Popularity
Time



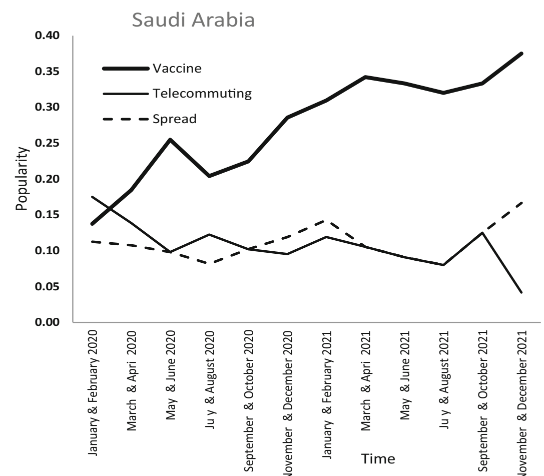
Chile
Vaccine
Death
Telecommuting
Popularity
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time



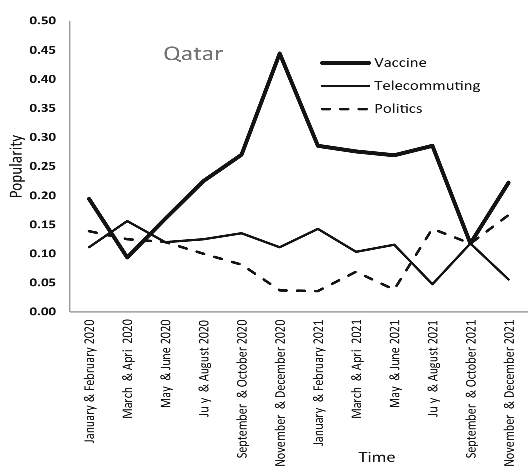
Denmark
Vaccine
Masking
Politics
Popularity
0.50
0.40
0.30
0.20
0.10
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time



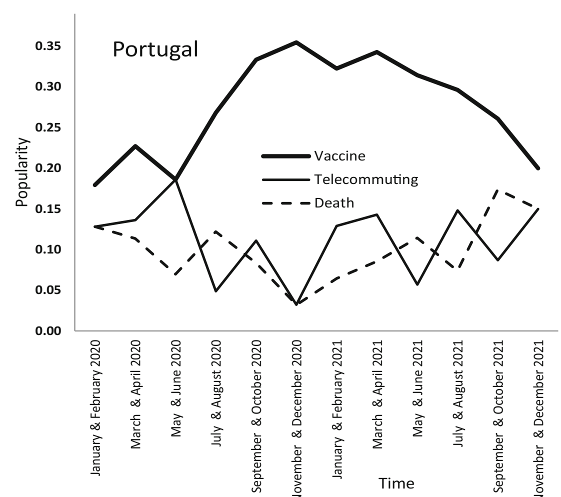
Peru
Vaccine
Death
Telecommuting
Popularity
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time



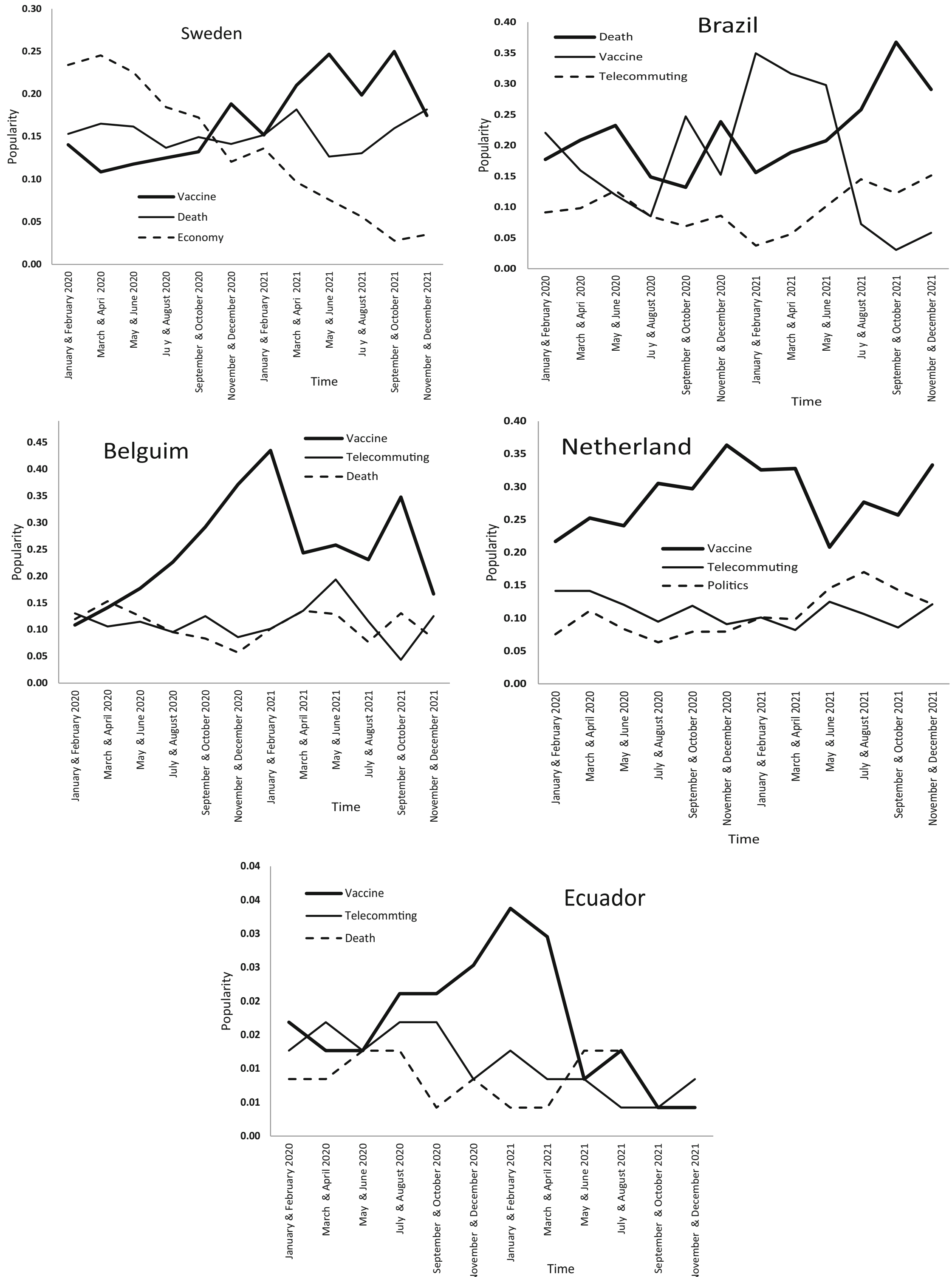
Saudi Arabia
Vaccine
Telecommuting
Spread
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time



Qatar
Vaccine
Telecommuting
Politics
Popularity
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time



Portugal
Vaccine
Telecommuting
Death
Popularity
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time



Sweden
Vaccine
Death
Economy
Popularity
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time
Brazil
Death
Vaccine
Telecommuting
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & Apri 2020
May & June 2020
Ju y & August 2020
September & October 2020
November & December 2020
January & February 2021
March & Apri 2021
May & June 2021
Ju y & August 2021
September & October 2021
November & December 2021
Time
Belguim
Vaccine
Telecommuting
Death
Popularity
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time
Netherland
Vaccine
Telecommuting
Politics
Popularity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time
Ecuador
Vaccine
Telecommting
Death
Popularity
0.04
0.04
0.03
0.03
0.02
0.02
0.01
0.01
0.00
January & February 2020
March & April 2020
May & June 2020
July & August 2020
September & October 2020
November & December 2020
January & February 2021
March & April 2021
May & June 2021
July & August 2021
September & October 2021
November & December 2021
Time

Author contributions Faezeh Azizi performed methodology, software, and original—draft preparation. Hamed Vahdat-Nejad conducted supervision, conceptualization, methodology, reviewing, and editing. Hamideh Hajiabadi provided supervision, reviewing, and editing.

Data availability https://github.com/FaezehAzizi1995/Paper3.git

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval Not applicable.

References

1. Blei, D.M.: Probabilistic topic models. Communications Assoc. Comput. Mach. **55**, 77–84 (2012)
2. Sun, L., Yin, Y.: Discovering themes and trends in transportation research using topic modeling. Transp. Res. Part C Emerg. Technol. **77**, 49–66 (2017)
3. Azizi, F., Hajiabadi, H., Vahdat-Nejad, H., Khosravi, M.H.: Detecting and analyzing topics of massive COVID-19 related tweets for various countries. Comput. Electr. Eng. **106**, 1–11 (2023)
4. Palanichamy, Y., Kargar, M., Zolfagharinia, H.: Unearthing trends in environmental science and engineering research: insights from a probabilistic topic modeling literature analysis. J. Clean. Prod. **317**, 1–21 (2021)
5. Xie, Y., Ning, C., Sun, L.: The twenty-first century of structural engineering research: a topic modeling approach. Structures **35**, 577–590 (2022)
6. Davies, D.L., Bouldin, D.W.: A cluster separation measure. Inst. Electr. Electron. Eng. Trans. Pattern Anal. Mach. Intell. **1**, 224–227 (1979)
7. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
8. Caliński, T., Harabasz, J.: A dendrite method for cluster analysis. Commun. Stat. Theory Methods **3**, 1–27 (1974)
9. Dunn, J.C.: Well-separated clusters and optimal fuzzy partitions. J. Cybern. **4**, 95–104 (1974)
10. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
11. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. **3**, 32–57 (1973)
12. Vickery, A., Vickery, B.C.: Information science in theory and practice. J. Doc. **61**, 814–815 (2005)
13. Doersch, C.: Tutorial on variational autoencoders", arXiv:1606.05908 (2016)
14. Srivastava, A., & Sutton, C.: Autoencoding variational inference for topic models. In: Fifth learning representations, pp. 1–12. International conference on learning representations (2016)
15. Fei, H., Ji, D., Zhang, Y., Ren, Y.: Topic-enhanced capsule network for multi-label emotion classification. Inst. Electr. Electron. Eng. Assoc. Comput. Mach. Trans. Audio Speech Lang. Process. **28**, 1839–1848 (2020)
16. Fei, H., Zhang, Y., Ren, Y., & Ji, D.: Latent emotion memory for multi-label emotion classification. In: Thirty-fourth association for the advancement of artificial Intelligence, USA, pp. 7692–7699. Association for the advancement of artificial Intelligence Press (2020)
17. Fei, H., Ren, Y., Wu, S., Li, B., Ji, D.: Latent target-opinion as prior for document-level sentiment classification: A variational approach from fine-grained perspective. In: Thirtieth international world wide web USA, pp. 553–564. Association for computing machinery, 2021
18. Fei, H., et al.: Lasuie: Unifying information extraction with latent adaptive structure-aware generative language model. Adv. Neural. Inf. Process. Syst. **35**, 15460–15475 (2022)
19. Fei, H., Wu, S., Zhang, M., Ren, Y. Ji, D.: Conversational semantic role labeling with predicate-oriented latent graph. arXiv:2210.03037. (2022)
20. Callaghan, M.W., Minx, J.C., Forster, P.M.: A topography of climate change research. Nat. Clim. Chang. **10**, 118–123 (2020)
21. Yu, J., Lu, Y., Muñoz-Justicia, J.: Analyzing Spanish news frames on Twitter during COVID-19—a network study of El País and El Mundo. Int. J. Environ. Res. Public Health **17**, 1–12 (2020)
22. Mahanty, S., Boons, F., Handl, J., Batista-Navarro, R.: An investigation of academic perspectives on the 'circular economy'using text mining and a Delphi study. J. Clean. Prod. **319**, 1–15 (2021)
23. Fang, D., Yang, H., Gao, B., Li, X.: Discovering research topics from library electronic references using latent Dirichlet allocation. Library Hi Tech **36**, 400–410 (2018)
24. Newman, D., Hagedorn, K., Chemudugunta, C., Smyth, P.: Subject metadata enrichment using statistical topic models. In: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries pp. 366-375. (2007)
25. Rani, M., Dhar, A.K., Vyas, O.: Semi-automatic terminology ontology learning based on topic modeling. Eng. Appl. Artif. Intell. **63**, 108–125 (2017)
26. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
27. McInnes, L., Healy, J., Melville, J.: Umap: uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 (2018)
28. Espadoto, M., Martins, R.M., Kerren, A., Hirata, N.S., Telea, A.C.: Toward a quantitative survey of dimension reduction techniques. Inst. Electr. Electron. Eng. Trans. Vis. Comput. Graph. **27**, 2153–2173 (2019)
29. Estivill-Castro, V.: Why so many clustering algorithms: a position paper. Assoc. Comput. Mach. Spec. Interest Group Knowl. Discov. Data Mini. Explor. Newslett. **4**, 65–75 (2002)
30. MacQueen J.: "Classification and analysis of multivariate observations", presented at the Fifth berkeley symposium on mathematical statistics and probability. (1967).
31. Murtagh, F.: A survey of recent advances in hierarchical clustering algorithms. Comput. J. **26**, 354–359 (1983)
32. Ezugwu, A.E., et al.: A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Eng. Appl. Artif. Intell. **110**, 1–43 (2022)
33. Castellanos, A., Cigarrán, J., García-Serrano, A.: Formal concept analysis for topic detection: a clustering quality experimental analysis. Inf. Syst. **66**, 24–42 (2017)