COVID-19-Related Research in CS Area: A Topic Analysis Approach

Yuxin Shao
ID 260853676

1 Introduction

The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing global pandemic of coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. As of mid-June 2021, there were 174 million confirmed cases worldwide [2]. To help to solve the different problems caused by the COVID-19 pandemic, researchers from a variety of fields are working hard to identify theoretically and practically useful solutions. For example, in the clinical practice guide field, [3] give a study of the importance of social distancing for COVID-19 infection likelihood. In the outbreak detection and analysis field, [4] give how contact patterns change changes the COVID-19 outbreak in China. In the topic analysis and social network field, [5] give an overall topic analysis of media news coverage in the early COVID-19 pandemic.

It is important to get a big picture of COVID-19 literature. On the one hand, this will give researchers a better understanding of what are the hot topics in the research field. On the other hand, this also helps researchers understand how different research fields interact with each other to solve the problems during the COVID-19 period. By focusing on the field of computer science (CS), we can also give researchers who are in other fields some information about how CS techniques can apply to their research fields and help them to solve their problems. To get a big picture of COVID-19-related researches in the CS area, we propose a topic-based analysis of COVID-19 related CS papers using a large database of scholarly publications. In particular, we answer the following research questions:

- Which topics have computer science researchers been most interested in throughout the COVID-19 pandemic?
- Is there any relation between the countries of affiliations for the published paper? What's the collaboration pattern between different countries, such as some countries tend to collaborate more with some specific countries? Are there any country-specific patterns, such as focus on specific topics?
- Is there a clear relation among some specific topics, such as co-occurrence pattern among topics? What does this phenomenon infer? What's the top co-occurred topics?

The rest of this work is organized as follows. Section 2 gives a review of the related work of this work including topic-based bibliometric analysis, topic modeling, and keyword network analysis. Section 3 gives a detailed methodology of the experiment, including data collection, topic modeling of the data, and keyword network formation and analysis. Section 4 gives the result of the experiment and an analysis of the result. Section 5 gives the conclusion of the result and analysis.

2 Related Work

2.1 Topic-based Bibliometric Analysis

Bibliometrics is the use of statistical methods to analyze books, articles, and other publications [6]. Bibliometric analysis is an effective way to measure the influence of publications, scholars, or institutions in the scientific community [7]. Topic-based bibliometric analysis is the analysis that

focuses on the topics extracted from the text and does a quantitative analysis on the publications. Most topic-based bibliometric analysis papers include topic modeling, topic focus analysis, topic change trending analysis, and analyzing the relationship between topics [8, 9, 10, 11]. Some of the papers also give a prediction for the future topic change trending [8].

Based on 6392 early COVID-19 related research publications selected from the COVID-19 Open Research Dataset (CORD-19), [12] applied a bibliometrics and topic analysis to analyze the fields of publications, citation behavior, and to find the topic distribution, collaborations, and innovative topics over time. Results indicate (1) research focus shift from pure Medicine and Biology field to mixing with social, economic, and psychological perspectives, (2) research topics develop interrelating with each other, (3) COVID-19 cases are highly correlated with the SARS pandemic, (4) research mainly focus on clinical characteristics and virus detection, (5) computer science is important for the research. These results give insights on the topic development and the evolution of COVID-19 research in the early stage. However, some aspects may need to be enhanced or modified to solve our questions. Firstly, as JiayingLiu et al. mentioned in the conclusion section of the paper, the dataset they used may be biased or cannot represent the COVID-19 related research publications. Secondly, as the pandemic is still ongoing, the result may need to be updated. Thirdly, because JiayingLiu et al. did a bibliometrics analysis on publications of all the research fields, there may be some detailed information of each single research field that cannot be captured.

Based on these aspects, we proposed a modified methodology that uses topic modeling for topic extraction from the publications, keyword network to find the relationship between topics, and focuses on the computer science field publications in 2020.

2.2 Topic Modeling

Topic modeling is a frequently used text-mining tool for the discovery of hidden semantic structures in a text body [13]. For topic-based bibliometric analysis, topic modeling is a very important procedure for the analysis. Existing topic modeling approaches can be divided into two main categories: probabilistic models and spectral methods [14]. For spectral methods, suggesting an algebraic recovery perspective and utilizes non-negative matrix factorization (NMF) as the main technique [14]. For probabilistic models, Probabilistic Latent Semantic Analysis/Probabilistic Latent Semantic Indexing (pLSA/pLSI) [15] and Latent Dirichlet Allocation (LDA) [16] are two examples.

NMF is a group of algorithms in multivariate analysis and linear algebra where a matrix V is factorized into two matrices W and H, with the property that all three matrices have no negative elements [17]. In mathematical format is $V \approx W \times H$. In the topic modeling context, V stands for documents-by-terms matrix, W stands for documents-by-topics matrix, and H stands for topics-by-terms matrix.

pLSA/pLSI is the improvement of Latent Semantic Analysis/Latent Semantic Indexing (LSA/LSI) [18]. LSA/LSI uses singular value decomposition (SVD) to reduce the number of rows of the tf-idf matrix while preserving the similarity structure among columns [19]. Unlike LSA/LSI using dimensionality reduction methods, pLSA/pLSI focuses more on probabilistic modeling to get the same effect of LSA/LSI.

Latent Dirichlet Allocation (LDA) is the most widely and frequently used method. The LDA model consists of a word set W, a document set D (corpus), and a topic set Z. A document is present by a word sequence $W = (w_1, ..., w_m)$, where the subscript is the index of word in the word sequence. A corpus is a set of documents where containing M documents denoted by $D = \{W_1, \ldots, W_M\}$. And the topic set is what we want to mine from the text. The number of the topic is defined by the user which is denoted by K. So, $Z = \{z_1, \ldots, z_k, ..., z_K\}$. Each topic z_k is determined by the conditional distribution probability $p(w \mid z_k)$ of a word $w \in W$. Each document W_m is determined by the conditional probability distribution $p(z \mid W_m)$ of a topic $z \in Z$.[14] There are two hyper-parameters α and β, controlling the amount of smoothing applied to the topic distributions for each document and the word distributions for each topic [12]. Gibbs sampling is also very frequently used in LDA instead of a variational Bayesian algorithm [16] to improve the accuracy of the posterior distribution.

Recently, with increasing researches in the AI field, some deep-learning-based topic modeling approaches are flourishing. Unlike the traditional topic modeling methods as we mentioned previously, deep-learning-based methods give more flexibility and scalability on the topic modeling process [20]. For example, the dynamic embedded topic model (D-ETM) proposed by [21] combines dynamic latent Dirichlet allocation (D-LDA) and word embeddings. By using structured amortized

variational inference with a recurrent neural network (RNN), [21] found that D-ETM outperforms D-LDA on document completion task with three different corpora. Besides, the researchers also found that D-ETM requires significantly less time to fit. Some deep-learning-based methods using other frameworks such as graph neural network (GNN), generative adversarial networks (GANs), autoregressive models, and so on [20].

In the study of COVID-19-related data, topic modeling is a very significant method. For example, [5] uses LDA as the topic modeling method to extract topics from traditional and social media for doing the topic analysis. [12] uses LDA to extract topics from the COVID-19-related research paper database and study the topic evolution and change over time. [22] uses Hawkes binomial topic model to get dynamic topics from the COVID-19 Twitter narrative among U.S. governors and cabinet executives. To get a stable and reliable result, we decide to use LDA, the most widely used method, as our topic modeling method.

2.3 Keyword Network Analysis

A keyword network is an undirected graph in which nodes are frequent or important words (keywords) found in a given corpus. There is an edge between two nodes if the two keywords are co-occurrence in an article. The main goal of keyword network analysis is to find information among keywords by utilizing the network structure. The edge weight is the frequency of words occurring together [23]. Thus the keyword network is undirected and weighted [24]. We can apply some well-known social network analysis methods to the keyword network.

Topology features analysis is the most natural one based on the network format. [25] used the weighted degree to analyze the hot keywords. They also use the weights of the edges to measure the strength of the co-occurrence of any two keywords. We can also take a look at degree centrality, betweenness centrality, characteristic path length, clustering coefficient, and density to get more information from the network [26].

Visualization of keyword networks is also very efficient for analysis and present the data. Many libraries or software, such as LaNet-vi[1] and Gephi[2], can visualize the keyword network and make the analysis clearer and more intuitive. Furthermore, we can group nodes in the keyword network into clusters. Based on the information provided by the keyword network, we can choose many algorithms for the clustering: hierarchy-based algorithms, information-theoretic algorithms, modularity-based community detection, and some other algorithms such as Network Community Structure Clustering Algorithm [27]. Clustering can help us better understand the relationship between keywords.

Inspiring by the keyword network analysis, we propose a method that using topics instead of keywords to build a "topic network" and doing the analysis on the network. By using this method, we can get a clear picture of the relationships among topics.

3 Methodology

3.1 Data Source

In this project, we use Scopus[3] as the data source to find the interested paper. Scopus is the largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings in the fields of science, technology, medicine, social sciences, and arts and humanities [28]. It supports using code to do the advanced search and exporting metadata for the result papers.

3.1.1 Data Collection

By using the query shown in Figure 1 to perform an advanced search on Scopus, we retrieve the 1560 articles and 2342 conference papers written in English that contain the phrases *"covid-19"*, *"sars-cov-2"*, or *"2019-ncov"* in the title or the abstract and were published in 2020. Because we decided to use article title and abstract as the source of topic modeling, we then use the export tool provided by Scopus to export the title and abstract of each article and conference paper together.

[1] `https://lanet-vi.fi.uba.ar/`

[2] `https://gephi.org/`

[3] `https://www.scopus.com/`

After retrieving the raw data, we de-duplicate the publications and remove those missing abstracts. We are then left with 1520 articles and 2332 conference articles, for a total of 3852 publications.

```
TITLE-ABS ( covid-19 OR sars-cov-2 OR 2019-ncov ) AND SUBJAREA ( comp ) AND
 DOCTYPE ( ar OR cp ) AND ( LIMIT-TO ( PUBYEAR , 2020 ) ) AND ( LIMIT-TO (
                          LANGUAGE , "English" ) )
```

Figure 1: The Scopus query used to obtain our dataset.

3.1.2 Data Pre-processing

After cleaning, we do the data pre-processing for the LDA which includes tokenization, stop-words removing, lemmatization, convert data into the bag-of-words format and removing too common and too rare terms.

Tokenization is used to split the text content of the title and abstract into lowercase separated words. Because the LDA model uses the document-term matrix as input, we need to first tokenize the raw text into terms for further manipulations. After that, we use the pre-defined stop-word dataset from the `nltk` python library[4] to remove all the stop-words from the result of tokenization. This will further remove the distracting terms from the input of the LDA model. Lemmatization is the algorithmic process that determines the lemma of a word based on its intended meaning [29]. This will help users to turn a word with different verification into one single word which will reduce the size of terms and also remove redundancy in the terms data. The bag-of-words model is a format of terms representation. We keep the distinct terms and record the occurrence of each term in each document. Based on the bag-of-words model, we can use the recorded occurrence to filter too common and too rare terms. We decide to filter out the term which occurs in less than 5 documents and more than 80% of the corpus size which is 3062.

After all these pre-processing processes, we can move to the topic modeling part.

3.2 Bibliometric Analysis

The bibliometric analysis can be divided into two different types which are performance analysis and science mapping. Based on the data we collect, we conduct the publication source analysis and affiliations analysis for both institutions and the country or country they belong to for the performance analysis part. And we conduct scientific collaboration analysis including authors and their institution analysis for the science mapping part.

Publication source analysis is the analysis based on the statistics data of the publication source of the articles in the dataset. Affiliations analysis is the analysis based on the statistics data of productive institutions and productive countries/countries which is including article count, citation count, and the ranking of each affiliation. Scientific collaboration analysis is for discovering the relation between different authors and their institutions. Usually, we can use a co-author and co-institution network to visualize the relation.

By doing bibliometric analysis, we can have a better overview idea of our articles dataset which helps us to solve the questions we ask at the beginning.

3.3 Topic Modelling: LDA

3.3.1 Hyper-parameter Selection

There are 3 hyper-parameter we can tune in the LDA model: k, α, and β where k is the topic number of the LDA topic model, α is the document-topic density, and β word-topic density.

For α, and β, we use the practical value setting of α, and β in our LDA model which was used by Griffiths TL et al. in 2004.[30] α is set to $50/k$ and β is set to 0.1. For topic number k, we examine the topic coherence value of the LDA model with different k values in the range [15, 20, 25, 30, 35, 40, 45, 50] to decide which value of k is the best for our dataset.

[4] `https://www.nltk.org`

3.3.2 Evaluation of LDA Model

To evaluation the LDA model, usually we use the perplexity and the topic coherence value. The perplexity, used by convention in language modeling, is algebraically equivalent to the inverse of the geometric mean per-word likelihood. A lower perplexity score indicates better generalization performance [16]. And topic coherence measures score a single topic by measuring the degree of semantic similarity between high scoring words in the topic. These measurements help distinguish between topics that are semantically interpretable topics and topics that are artifacts of statistical inference [31]. A higher topic coherence score indicates a better topic modeling performance. As perplexity doesn't consider the context and semantic associations between words, a good topic number measured by perplexity standard may lead to the artificial topic result. So we use topic coherence here to evaluate our LDA model and tune the hyper-parameters.

3.3.3 Labeling Topics

After getting the topic result from the LDA model, we manually label the topics with the human-interpretable topic title based on the dominant words each topic contains and the professional knowledge we have. This helps us to do further analysis in the next part of the experiment.

3.4 Topic Network Analysis

3.4.1 Topic Co-occurrence Network Construction

Given the result of topic modeling, we assign the two most related topics to each article. Using topics as the node, creating an edge between two topics if they are assigned to the same article, and using weights of the edge to represent the connection number. We can derive this relationship by a co-occurrence matrix of topics.

3.4.2 Network Visualization

After determining the structure of the network, we use NetworkX[5] and Gephi to visualize the network. NetworkX is a python library which dedicates to manipulate and visualize network relating work and Gephi is a software which can visualize network in a better-looking way. Visualization can help us know the relationship between topics better and clear.

3.4.3 Topological Features Analysis

There are a lot of measurements we can do on the topic network. Degree, clustering coefficient, shortest path, degree centrality, and so on.

The degree is the number of connections a node has. As we use a weighted network, we need to consider the weights as well. We use the sum of weights of all edges of a node to calculate the weighted degree. The degree is often used to measure the strength of each node's connection to other nodes [25]. Local clustering coefficient is the measure of connectivity of the nodes and their neighbors and average clustering coefficient is the measure of overall connectivity of the network. The shortest path is the shortest distance between two nodes. We usually use this to imply the connectivity of the network and how information flow in the network. Degree centrality is defined as the number of links incident upon a node [32]. Because we use an undirected network, the degree centrality for a node is simply its degree [33].

By calculating these topological features, we can better understand the network structure and implicit information in the network. After that, we do the analysis based on the topological features we get.

3.4.4 Clustering Analysis

To study the implicit grouping in the topic network, we use the python implementation of Louvain [34], a heuristic method that is based on modularity optimization [35]. We then visualize the result of clustering and do further analysis on different clusters.

[5] `https://networkx.org`

4 Result and Analysis

4.1 Bibliometric Analysis

4.1.1 Countries Collaboration Analysis

We study the country each paper's affiliations belong to and try to get the collaboration pattern among different countries. We sorted the countries by the number of papers they published and filter out the countries with the number of published papers greater than 50, see Figure 2. United States, China,

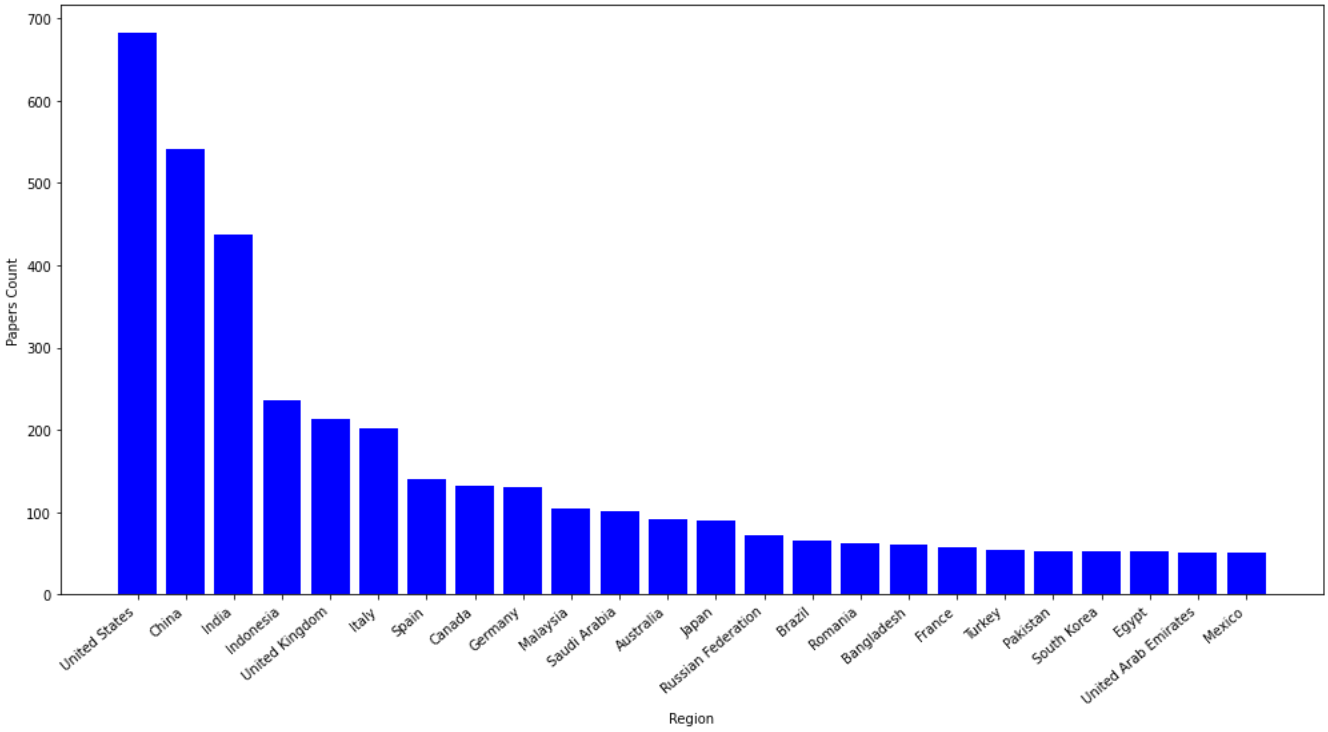


Figure 2: Published papers count for each country with count > 50

India, Indonesia, United Kingdom are the top five prolific countries where the United States, China, and India have significantly higher total published papers count than the rest of the countries.

Comes to collaboration rate, France, Pakistan, Saudi Arabia, United Kingdom, United Arab Emirates, and South Korea are the top five countries. Japan, China, Romania, Russian Federation, India, and Indonesia are the least five countries, see Figure 3.

To further study the collaboration pattern, for each country, we study the collaboration proportion with other countries (see Appendix A). We can find that United States, India, and China have the most number of collaboration countries, which is 22. Russian Federation has the least number of collaboration countries, which is 9. We focus on these countries with extreme value. The United States and China have each other as the top collaborated country. India has the United States as the top collaborated country and the United States has India as the second collaborated country. However, China and India do not collaborate as much as the previous relationship does. China and India collaboration only occupy 1.6% in China collaboration distribution which is the 14th place. In India's distribution, it is 2.7% which is the 8th place. Russian Federation also has United State as the top collaborated country and has China, Spain, and France as the second, India as the third.

We also build a country collaboration network (Figure 4) to study the overall pattern of inter-countries collaboration. The network consists of countries as nodes and edges represent the collaboration between countries. The average degree of the network is 16.33 which means the average number of collaborated countries per country is 16.33. The graph density is 0.71. The density of a graph is a measure of how many ties between actors exist compared to how many ties between actors are

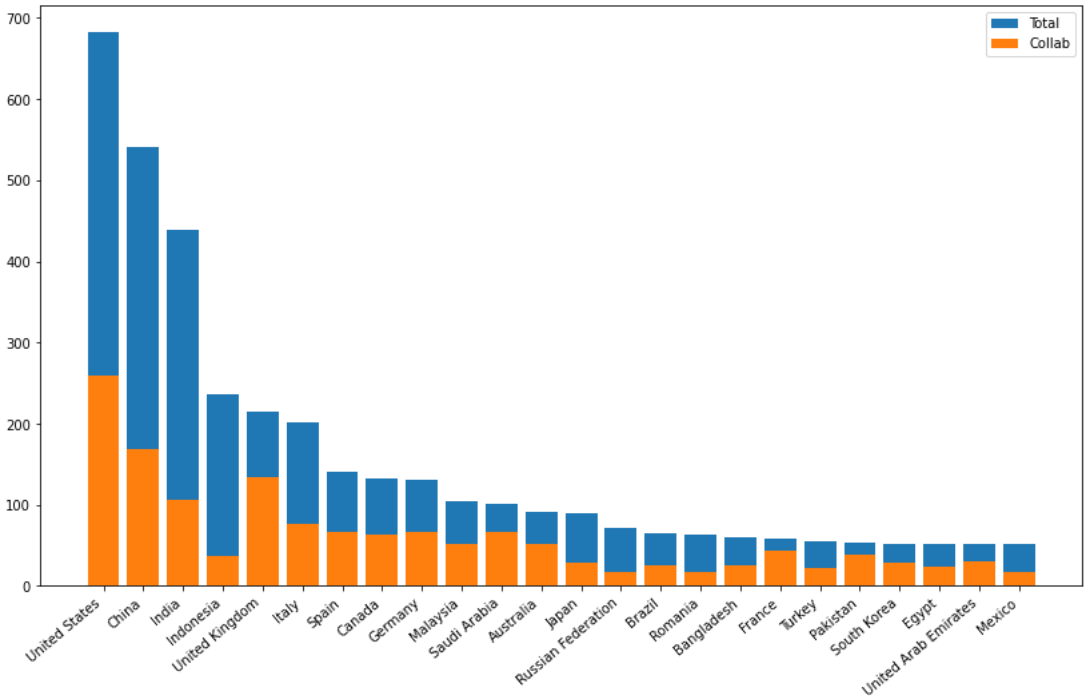


Figure 3: Total and collaboration count for country with count of published papers > 50

possible [36]. 0.71 shows that this network is not fully connected and with a high density. This means that countries collaborate at a high rate. From the network visualization, we can find that commonly, a country with a small node size will have a darker node color, such as France and Pakistan. This shows that countries with a smaller number of published papers will tend to collaborate with other countries more. On the other hand, a country with a bigger node size will have a lighter node color, such as United States, China, and India.

4.1.2 Bibliographic Coupling Analysis

Bibliographic coupling is a technique for science mapping that operates on the assumption that two publications sharing common references are also similar in their content [37]. We build a co-reference network using papers as nodes to study bibliographic coupling. If there is a co-occurrence of reference between two papers, we add an edge. The edges are weighted by the number of occurrences of references between two papers.

After building the network, we find that the highest edge weight between two nodes is 54, which means that there are 54 common references between two papers in the dataset. After removing 1367 nodes which are isolated to the rest of the graph, we get 2485 nodes left. The average degree is 77.96, which is how many other papers a given paper shares at least one reference with on average in the filtered dataset. The graph density is 0.031 which is very low. We then perform the Louvain method for community detection on this new graph and get 22 clusters in the graph (Figure 5).

Because of the assumption that two publications sharing common references are also similar in their content, we have the reason to say that the clusters inferred in the network may imply the topics among the papers. We try to use keyword frequency within the cluster to give our topic label for each cluster. After calculating the frequency of each keyword in the same cluster, we get the top five most frequently occur keywords in the cluster and we use the keyword to label the cluster topic. See the top keywords result for each cluster in Appendix B. Because only clusters 0,1,2,3,6,8 have considerably higher nodes numbers, we only focus on these clusters here. From the top 5 keywords we get, we have the following inferred topic labels Table 1.

We will discuss further about this later in the section 4.2 by comparing the result of LDA model and the result of co-reference network clustering.

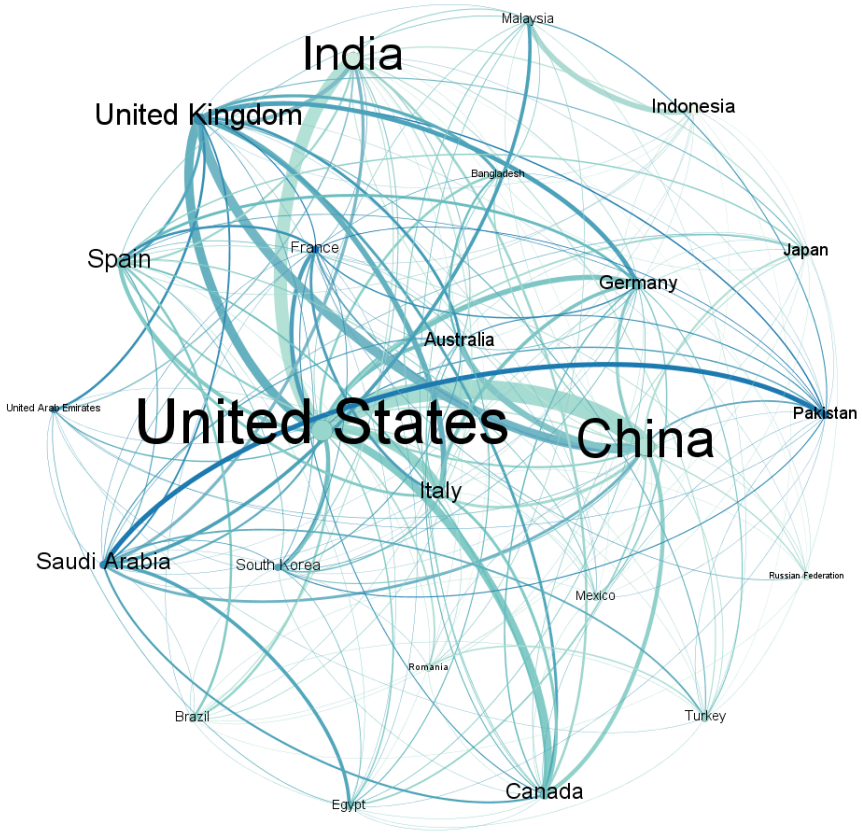


Figure 4: country collaboration network ranking node size with #published papers, ranking node color with collaboration rate, ranking label with degree

4.1.3 Co-authorship Analysis

Co-authorship analysis examines the collaborations among scholars in a research field. We use the data we collect to build a co-authorship network, where nodes represent the author and edges represent the co-authorship. After filtering out the authors with less than 2 papers published, and visualize the network, we can see the pattern of the collaboration among authors (Figure 6). Most of the collaborations happen within a certain sub-network, which means that some authors work together and do not collaborate outside a certain community. However, there are a few exceptions that connect two sub-graphs together, such as the one in Figure 7. The connection node in this sub-network represent author Abdulkareem, Karrar Hameed and he produces 3 papers with the authors within this sub-network which are *Helping doctors hasten COVID-19 treatment: Towards a rescue framework for the transfusion of best convalescent plasma to the most critical patients based on biological requirements via ml and novel MCDM methods*, *COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images*, and *Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods*. In these three papers, *COVID-CheXNet* and *Benchmarking Methodology* collaborate within the same sub-network and *Helping doctors* comes from the other sub-network. The previous two papers are more related to the pure computer science field with the medical context, but the latter one is more related to biology and the medical field using the computer science method. Abdulkareem connects two sub-networks that represent different research fields together and form a new sub-network.

Figure 5: Co-reference network partitions into clusters with color

There are other things we can notice from the co-authorship. Checking the network without filter out the author with less than 2 papers published, we find a paper with 86 authors by reversed sort the nodes in the network by degree. The title of the paper is *International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium*. Also, we sort the edges by the weight they have and find the top one is that two authors collaborate to published 10 papers together and both of them did not publish any other paper during 2020. The two authors are Anwesh Reddy Paduri and Narayana Darapaneni.

Cluster no.	#Nodes	Inferred topic
0	324	Image Detection
1	388	Covid-19 Drugs
2	577	Model Prediction
3	716	Online Learning
6	85	Preventing Measure
8	352	Social Media

Table 1: Inferred topic in each cluster of the co-reference network

4.2 Topic Modelling

4.2.1 Topics and Their Keywords

We choose 22 as the k value of LDA and get the 22 topics and their top 10 most significant words in the dataset, see Appendix C. We use these significant words within the topic model result, the most representative paper (the paper with highest percentage for a given topic) for each topic, and combined with our knowledge to infer the interpretable topic labels to each topic. The inferring topic labels see Table 2.

The "*" sign in Table 2 means that we cannot give one dominant topic label. The papers of these topics come from multiple big topics and cannot be given an overall interpretable topic label. For instance, we choose topic 1 to see the detail. After sorting the papers with topic 1 as the first dominant topic by the percentage of topic 1 occupied in each paper, we get the top 10 representative papers of topic 1 (Table 3). We can see that the top 10 representative topics can fall into different big topics, such as machine learning, IT, COVID-19 impact, and so on. Because the most representative paper of topic 1 has only 24.9% topic 1 occupation, we can infer that the papers in topic 1 cannot be well labeled under a certain interpretable topic label. Other topics with the "*" sign in Inferring Topic Label column have the same situation.

This problem may be due to the choosing of the k-value of our LDA model. Because we want to study more specific topics of the paper dataset, we choose a k-value with a second optimal topic coherence value (see Figure 8). We get the best topic coherence value around 7 topics and get a second peak at around 21 topics. This decision may introduce artificial topics to the result of the LDA model.

4.2.2 Topics of Papers

We choose 2 dominant topics to assign to each paper, here is the result of the first 10 papers in Table 4.

We also get the number of papers by the 2 dominant topics within the dataset (Figure 9). The total count of papers should be 2 times the dataset size as we assign 2 topics to each paper. As we can see from the bar graph, during 2020, researchers focused on topic 0: online education most. There are 764 papers in the dataset related to online education. This reflects the issue of in-person education cannot be done during the Covid-19 pandemic. Topic 15: image detection is the second most focused topic. This topic is related to using machine learning or deep learning technique to detect the image. In this topic, most papers are related to using image detection to help Covid-19 diagnosis. Topic 4 is about using modelling to help predict the Covid-19 pandemic trend in the future, which is also a very popular topic during 2020.

4.2.3 Comparing LDA Method and Co-reference Network Clustering

From section 4.1.2, we build a co-reference network to detect the implicit topics in the dataset by the assumption that two publications sharing common references are also similar in their content. Here, we do a comparison between the result of co-reference network clustering and the LDA model. We adding the topics to the co-reference network which removing the isolated nodes in section 4.1.2.

Topic	Inferring Topic Label	Most Representative Paper
0	Online Education	Implementation online lectures in Covid-19 Pandemic: A student perception
1	*	A deep learning based approach to child labour detection
2	*	Alignment of the marshall grazing incidence X-ray spectrometer (MaGIXS) telescope mirror and spectrometer optics assemblies
3	*	Action-entropy Approach to Modelling of 'Infodemic Pandemic' System on the COVID-19 Case
4	Model prediction	Coronavirus Outburst Prediction in India using SEIRD, Logistic Regression and ARIMA Model
5	Misinformation	Time series based trend analysis for hate speech in twitter during COVID 19 pandemic
6	Social Media	Loneliness, boredom and information anxiety on problematic use of social media during the COVID-19 pandemic
7	Epidemics Spread Control	Prevention and Control Strategy for Multi-group Epidemics Based on Delay and Isolation Control
8	Preventing Covid-19 infection	Human face recognition and temperature measurement based on deep learning for covid-19 quarantine checkpoint
9	*	A study on exercise recommendation method using Knowledge Graph for computer network course
10	*	Explaining factors affecting telework adoption in South African organisations pre-COVID-19
11	Data Analysis	Big Data Visualization and Visual Analytics of COVID-19 Data
12	Tracking System	Proximity Tracing in an Ecosystem of Surveillance Capitalism
13	Covid-19 Drugs	Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2
14	Security and Privacy	A secure and distributed framework for sharing COVID-19 patient reports using consortium blockchain and IPFS
15	Image Detection	Transfer Learning Based Method for COVID-19 Detection from Chest X-ray Images
16	Remote Life and Security	Home working and cyber security 2013 an outbreak of unpreparedness?
17	*	KIP Recipient Decision Making for Students Affected by Covid 19 Pendemi Using Fuzzy MADM Method
18	Economic Related	Supply Chain Finance for Targeted Poverty Alleviation: A Case Study of Suning
19	*	Fine-Grained Named Entity Recognition with Distant Supervision in COVID-19 Literature
20	AI in Healthcare	Building and Deploying a COVID-19 Monitoring Solution in March
21	Covid-19 Impact	Air quality during SARS-CoV-2 (COVID-19) lockdown in Sarajevo

Table 2: Inferring Topic Label

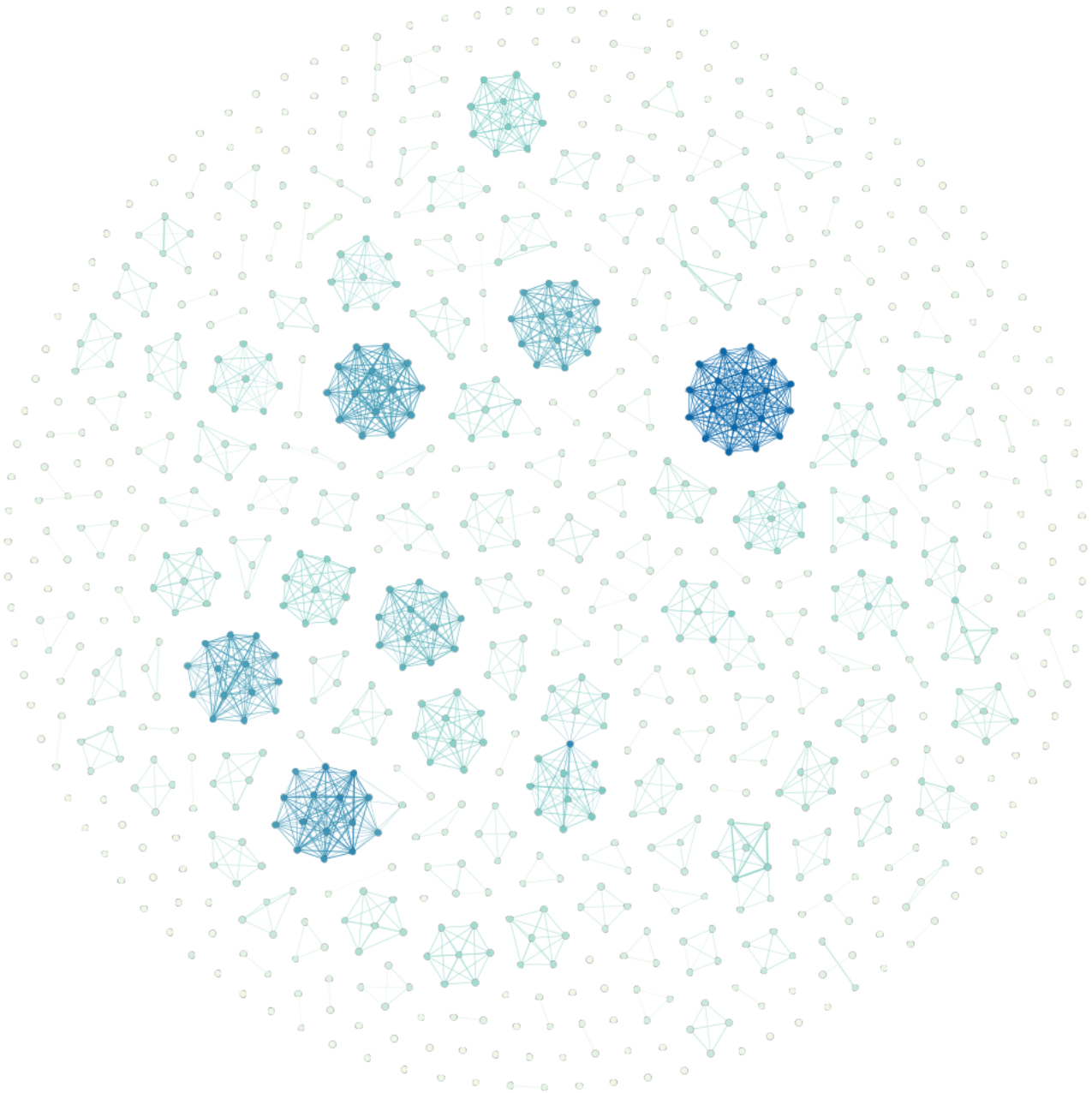


Figure 6: Co-author network colored by degree

Figure 10 is the result visualization. Comparing to Figure 5, which is the visualization with clustering result of the co-reference network, we can find some similar color clustering patterns, but also some of them do not match very well. We find all the papers which are clustering as 0 in the co-reference network, which is the light blue color in the visualization (Figure 5) and check if their topics labels are under the same topic label. We find that the top topic is topic 15 which is the dark green color in the visualization (Figure 10) and the match percentage is 76.23%. Because clusters that besides cluster 0,1,2,3,6,8 have very little nodes assign to them (below 10 nodes), we focus on cluster 0,1,2,3,6,8 in the following analysis. We conduct the above procedure to each cluster and get the result in Table 5.

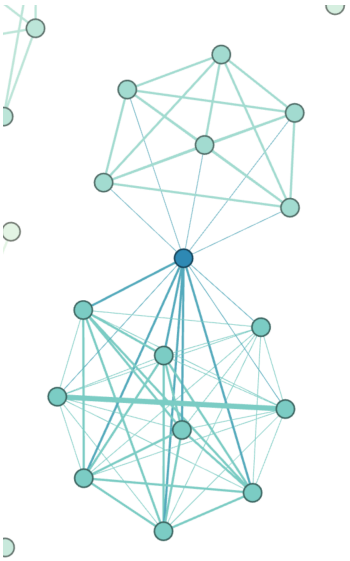


Figure 7: A sub-network in the previous co-author network

Title	**1st_topic_percentage**
A deep learning based approach to child labour detection	24.90%
Decision Support System for Selection of Staples Food and Food Commodity Price Prediction Post-COVID-19 Using Simple Additive Weighting and Multiple Linear Regression Methods	22.49%
COVID-19 Spread prediction Based on Food Categories using Data Science	20.60%
Judging pre-owned-bicycle deals type using machine learning methods	18.70%
Every document is born "grey"-Some documents can become "open"	17.67%
COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective	17.56%
Vision: A Critique of Immunity Passports and W3C Decentralized Identifiers	15.49%
Digital Transformation in Academic Society and Innovative Ecosystems in the World beyond Covid19-Pandemic with Using 7PS Model for IoT	15.35%
A comprehensive survey of IT sectors affected by covid-19	14.80%
Role of mobile communication with emerging technology in COVID'19	14.72%

Table 3: Top 10 representative papers of topic 1

For cluster 0, we get a very good match percentage between the LDA model and co-reference network clustering, which is 76.23%. For clusters 1, 2, and 6, we get a not bad match percentage, which is around 46%. But for clusters 3 and 8, the LDA model and co-reference network cluster match about 1/3 nodes in the cluster, but also differ in the rest 2/3 nodes. Further, when we comparing the label we get from the dominant topic and the cluster inferred topic label, we can see that the two topic labels getting from different methods are very similar to each other.

From above observation, we can conclude that clustering in the co-reference network does give us some insight into topics within the corpus which also matches our assumption that two publications sharing common references are also similar in their content.

Title	1st_topic	1st_topic_percentage	2nd_topic	2nd_topic_percentage
MCNN: A deep learning based rapid diagnosis method for COVID-19 from the X-ray images	15	29.78%	17	7.51%
French-language COVID-19 terminology International or localized?	5	13.18%	17	12.42%
Exploring working group's psychological subjectivity on public smart work services in a cloud-based social networking	16	16.42%	14	13.02%
A novel virtual screening procedure identifies Pralatrexate as inhibitor of SARS-CoV-2 RdRp and it reduces viral replication in vitro	13	41.29%	2	12.33%
Geospatial mapping, Epidemiological modelling, Statistical correlation and analysis of COVID-19 with Forest cover and Population in the districts of Tamil Nadu, India	4	25.83%	11	13.60%
The Use of UTAUT Model for Understanding Academicians' Perception towards Online Faculty Development Programs (FDP)	10	21.58%	0	14.68%
A Novel Medical Support Deep Learning Fusion Model for the Diagnosis of COVID-19	15	23.32%	17	9.88%
Coverage COVID 19 with E-Learning Replacement: Review Paper	0	14.33%	17	9.57%
Refute the Decision of Auto-Promotion and Real Facts of Digital Online Classes during the Pandemic in Bangladesh	0	18.02%	1	12.86%
Social Distance Alert System to Control Virus Spread using UWB RTLS in Corporate Environments	12	15.56%	8	13.23%

Table 4: Topics for papers

Cluster no.	#Nodes	Dominant Topic	Dominant Topic Label	Match Percentage	Cluster Inferred Topic Label
0	324	15	Image Detection	76.23%	Image Detection
1	388	13	Covid-19 Drugs	54.12%	Covid-19 Drugs
2	577	4	Model prediction	41.42%	Model Prediction
3	716	0	Online Education	33.1%	Online Learning
6	85	8	Preventing Covid-19 in-fection	41.18%	Preventing Measure
8	352	5	Misinformation	29.26%	Social Media

Table 5: Dominant topic in each cluster of the co-reference network

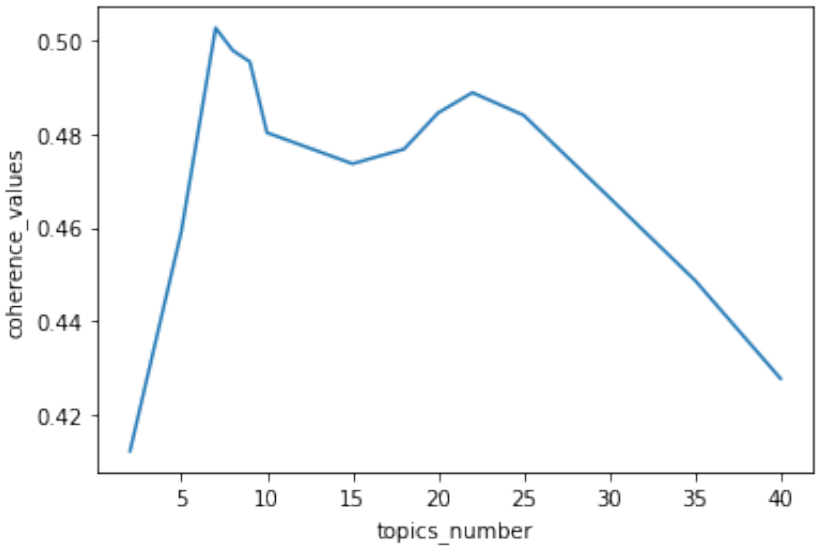


Figure 8: topic number vs. topic coherence value for LDA model

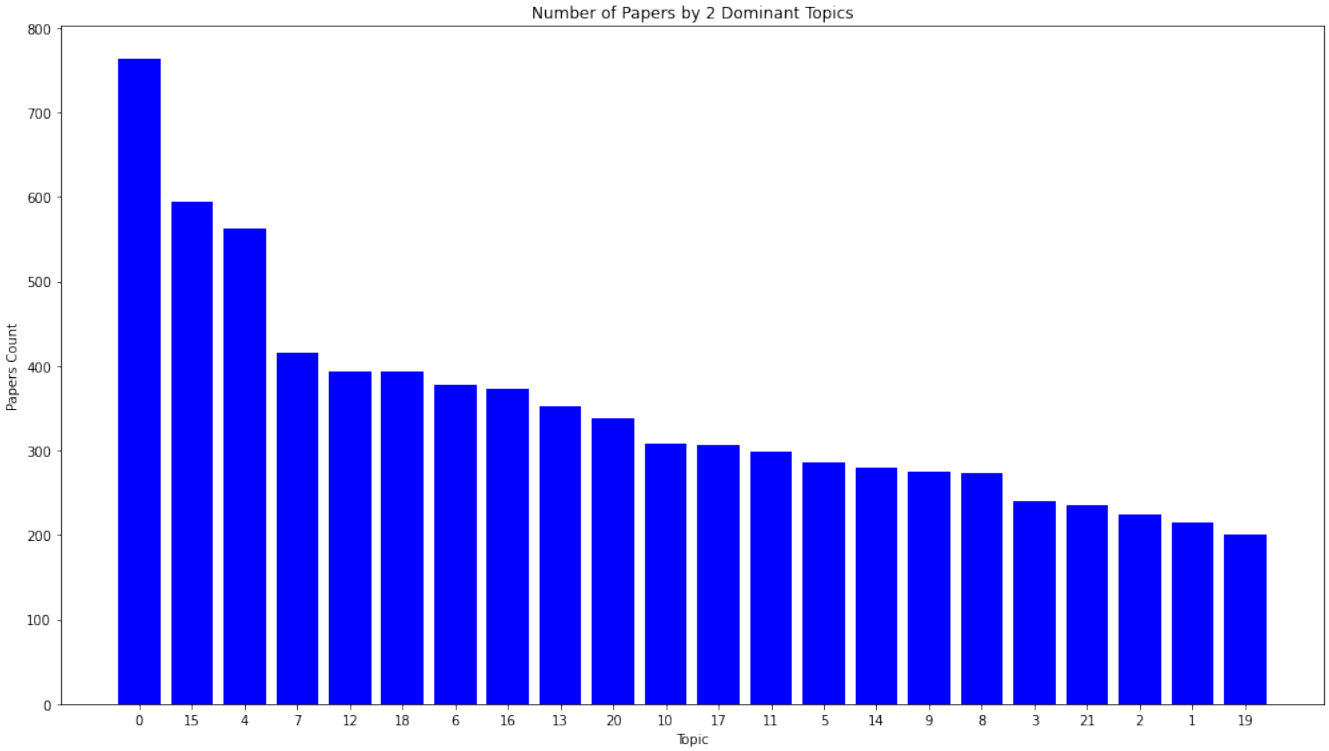


Figure 9: Number of Papers by 2 Dominant Topics

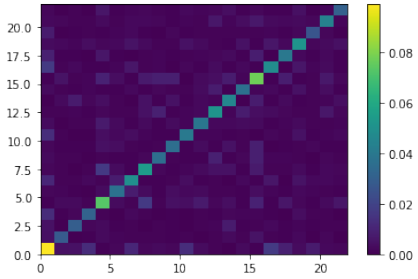
4.3 Topic Analysis

4.3.1 Topics Distribution within Each country

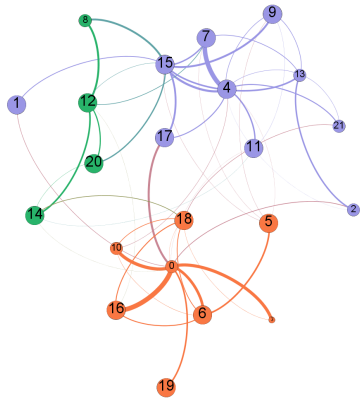
As we did the countries Collaboration Analysis in section 4.1.1, we also want to know if different countries have different research focuses during 2020 in cs field. So we categorize papers by the country of their affinities and find the topics distribution within each country. For the result, see Appendix D topics distribution within each country. Regarding the top topic for each country, topic

Figure 10: Co-reference network partitions into clusters with color by the result of LDA topic modelling

15 and topic 0 are the most popular topics. There are 10 countries with topic 15 as the most focused topic and 9 countries with topic 0 as the most focused topic. Some countries have fairly equal topic focus, such as United States, United Kingdom, but some other countries focus on some specific topics. Japan and Russian Federation focus more on topic 0 which is the online education topic and Turkey, Egypt, Mexico focus more on topic 15 which is the image detection topic. Besides, countries in the same continent may share the research focus. Japan, Russian Federation, Indonesia, Malaysia, and China share the most focused topic. In Europe, Italy, Germany, and France share the most focused topic. However, there are also some exceptions, such as South Korea in east Asia.



(a) Topics co-occurrence heat map



(b) Topic co-occurrence network with edges whose weight > 20

Figure 11: Topics co-occurrence visualizations

4.3.2 Topic Co-occurrence Analysis

We analysis the co-occurrence of each topic and visualized as a topic co-occurrence heat map and topics co-occurrence network in Figure 11. From the visualization, we can find that topic 7 Epidemics Spread Control and topic 4 Model prediction co-occur frequently in the corpus. Topic 0 Online Education and topic 16 Remote Life and Security also co-occur frequently. Using the clustering method in the topic co-occurrence network, we can find 3 clusters. Topic 0,10,18,16,6,19,5 group together as a cluster; topic 8,12,14,20 group together as a cluster; topic 1,2,4,7,9,11,13,15,17,21 group together as a cluster. Topics in a cluster have a higher co-occurrence weight than the other topics from other clusters. Topics with higher co-occurrence frequency tend to have some similarities in context thus tend to occur in the same paper.

5 Conclusion

The purpose of this report is to give a big picture of CS field research foci during the COVID-19 pandemic. By performing countries collaboration analysis, bibliographic coupling analysis, co-authorship analysis, and topic modeling, we get the following results.

- Researchers focus more on the topic related to online education during the pandemic time, image detection for Covid-19, and using models to predict how Covid-19 will spread in the future.
- United States, China, India are the top 3 prolific countries during 2020 in CS fields. The average collaboration rate of all countries whose published papers number is greater than 50 is 44.70%. The countries that have high published papers number tend to have lower collaboration rates. Some countries have fairly equal topic focus, such as United States, United Kingdom, but some other countries focus on some specific topics. countries in the same continent tend to share the same topic focus. However, there is no clear relationship between the geographic location of countries and the collaboration between countries.
- Researchers tend to collaborate in a certain community and most of the communities do not interact with each other.
- Topics tend to co-occur if they have some similar context. Epidemics spread control and model prediction, online education and remote life and security are two top co-occurrence topics.

Overall, this report provides readers an overview of the topics and collaboration pattern of papers about Covid-19 during 2020 in CS fields. It helps readers to understand how CS-related techniques can help to solve different fields' problems during the Covid-19 pandemic and how researchers and countries collaborate to solve the problems.

There are also some limitations in our report. We only use the data from Scopus and do not consider other sources. We also do not study the temporal trending of research focus change. In the future, we may focus on these limitations to do further research.

References

[1] Covid-19 pandemic, Jun 2021. URL https://en.wikipedia.org/wiki/COVID-19_pandemic.

[2] Health Organization World. Coronavirus disease (covid-19), 2020. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[3] Russell H. Fazio, Benjamin C. Ruisch, Courtney A. Moore, Javier A. Granados Samayoa, Shelby T. Boggs, and Jesse T. Ladanyi. Social distancing decreases an individual's likelihood of contracting covid-19, Feb 2021. URL https://www.pnas.org/content/118/8/e2023131118.

[4] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, and et al. Changes in contact patterns shape the dynamics of the covid-19 outbreak in china, Jun 2020. URL https://science.sciencemag.org/content/368/6498/1481.

[5] Chipidza W;Akbaripourdibazar E;Gwanzura T;Gatto NM;. Topic analysis of traditional and social media news coverage of the early covid-19 pandemic and implications for public health communication, Mar 2021. URL https://pubmed.ncbi.nlm.nih.gov/33653437/.

[6] Bibliometrics, May 2021. URL https://en.wikipedia.org/wiki/Bibliometrics.

[7] Bibliometric analysis. URL https://encyclopedia.pub/2024.

[8] Oliver Faust. Documenting and predicting topic changes in computers in biology and medicine: A bibliometric keyword analysis from 1990 to 2017, Mar 2018. URL https://www.sciencedirect.com/science/article/pii/S2352914818300534.

[9] Xieling Chen, Di Zou, Haoran Xie, and Fu Lee Wang. Past, present, and future of smart learning: a topic-based bibliometric analysis, Jan 2021. URL https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-020-00239-6.

[10] Xieling Chen Tianyong Hao. A topic-based bibliometric analysis of two decades of research on the application of technology in classroom dialogue - tianyong hao, xieling chen, yu song, 2020, Dec 2020. URL https://journals.sagepub.com/doi/abs/10.1177/0735633120940956.

[11] Xieling Chen, Di Zou, Haoran Xie, and Gary Cheng. A topic-based bibliometric review of computers in human behavior: Contributors, collaborations, and research topics, Apr 2021. URL https://www.mdpi.com/2071-1050/13/9/4859/htm.

[12] Jiaying Liu, Hansong Nie, Shihao Li, Xiangtai Chen, Huazhu Cao, Jing Ren, Ivan Lee, and Feng Xia. Tracing the pace of covid-19 research: Topic modeling and evolution, Apr 2021. URL https://www.sciencedirect.com/science/article/pii/S2214579621000538.

[13] Topic model, Jun 2021. URL https://en.wikipedia.org/wiki/Topic_model.

[14] Wu Wang, Houquan Zhou, Kun He, and John E. Hopcroft. Learning latent topics from the word co-occurrence network. In Dingzhu Du, Lian Li, En Zhu, and Kun He, editors, *Theoretical Computer Science*, pages 18–30, Singapore, 2017. Springer Singapore. ISBN 978-981-10-6893-5.

[15] Thomas Hofmann. Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99*, 1999. doi: 10.1145/312624.312649.

[16] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, Mar 2003. URL https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

[17] Non-negative matrix factorization, Sep 2021. URL https://en.wikipedia.org/wiki/Non-negative_matrix_factorization.

[18] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis, Jan 1999. URL https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199009%2941%3A6%3C391%3A%3AAID-ASI1%3E3.0.CO%3B2-9.

[19] Latent semantic analysis, Aug 2021. URL https://en.wikipedia.org/wiki/Latent_semantic_analysis.

[20] He Zhao, Dinh Phung, Viet Huynh, Yuan Jin, Lan Du, and Wray L. Buntine. Topic modelling meets deep neural networks: A survey. *CoRR*, abs/2103.00498, 2021. URL https://arxiv.org/abs/2103.00498.

[21] Adji B. Dieng, Francisco J. R. Ruiz, and David M. Blei. The dynamic embedded topic model. *ArXiv*, abs/1907.05545, 2019.

[22] Hao Sha, M. Hasan, G. Mohler, and P. Brantingham. Dynamic topic modeling of the covid-19 twitter narrative among u.s. governors and cabinet executives. *ArXiv*, abs/2004.11692, 2020.

[23] Muskan Garg and Mukesh Kumar. The structure of word co-occurrence network for microblogs, Aug 2018. URL https://www.sciencedirect.com/science/article/pii/S0378437118309361.

[24] Arjun Duvvuru, Sagar Kamarthi, and Sivarit Sultornsanee. Undercovering research trends: Network analysis of keywords in scholarly articles. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, pages 265–270, 2012. doi: 10.1109/JCSSE.2012.6261963.

[25] Huajiao Li, Haizhong An, Yue Wang, Jiachen Huang, and Xiangyun Gao. Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network, Jan 2016. URL https://www.sciencedirect.com/science/article/pii/S037843711600025X.

[26] Jinho Choi, Sangyoon Yi, and Kun Chang Lee. Analysis of keyword networks in mis research and implications for predicting knowledge evolution, Sep 2011. URL https://www.sciencedirect.com/science/article/pii/S0378720611000784.

[27] B. Saha, Amitabha Mandal, S. B. Tripathy, and D. Mukherjee. Complex networks, communities and clustering: A survey. *ArXiv*, abs/1503.06277, 2015.

[28] Elsevier. What is scopus about? URL https://service.elsevier.com/app/answers/detail/a_id/15100/supporthub/scopus/.

[29] Lemmatisation, May 2021. URL https://en.wikipedia.org/wiki/Lemmatisation.

[30] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics, Apr 2004. URL https://www.pnas.org/content/101/suppl_1/5228.

[31] Shashank Kapadia. Evaluate topic models: Latent dirichlet allocation (lda), Dec 2020. URL https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0.

[32] Deepak Sharma and Avadhesha Surolia. *Degree Centrality*, pages 558–558. Springer New York, New York, NY, 2013. ISBN 978-1-4419-9863-7. doi: 10.1007/978-1-4419-9863-7_935. URL https://doi.org/10.1007/978-1-4419-9863-7_935.

[33] Jennifer Golbeck. Analyzing networks, Mar 2015. URL https://www.sciencedirect.com/science/article/pii/B9780128016565000214.

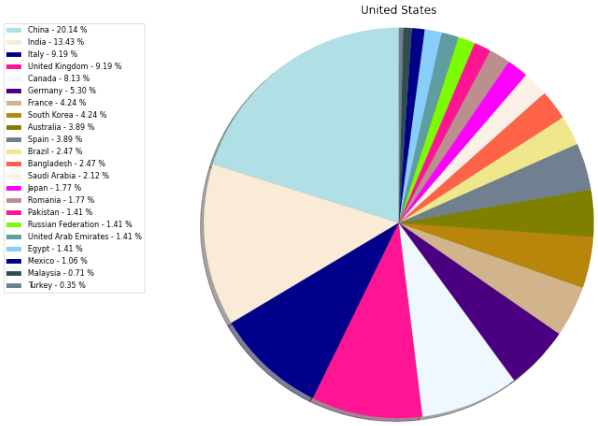
[34] Taynaud. taynaud/python-louvain. URL https://github.com/taynaud/python-louvain.

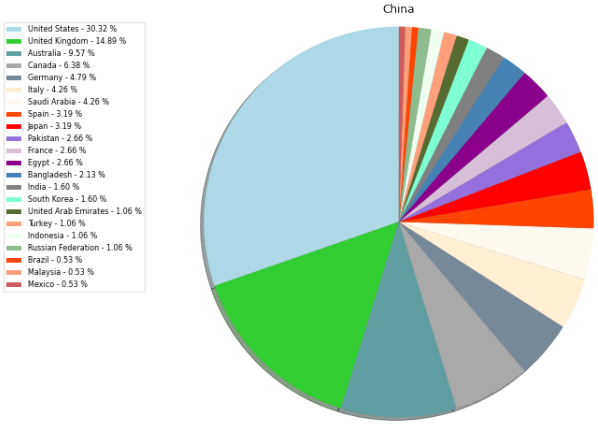
[35] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. URL https://ui.adsabs.harvard.edu/abs/2008JSMTE..10..008B/abstract.

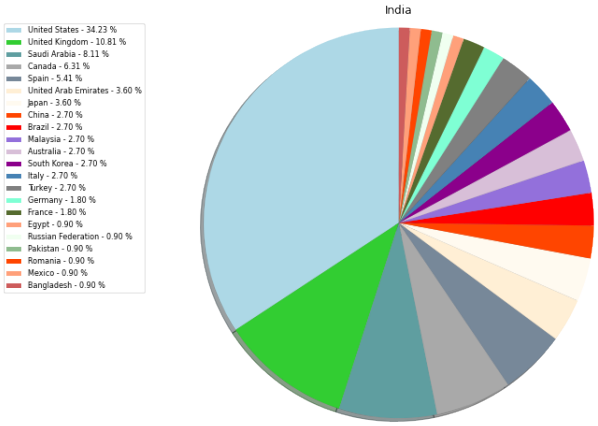
[36] Omar Lizardo Jilbert and Isaac. 2.9 density: Social networks: An introduction, Jan 2020. URL https://bookdown.org/omarlizardo/_main/2-9-density.html.

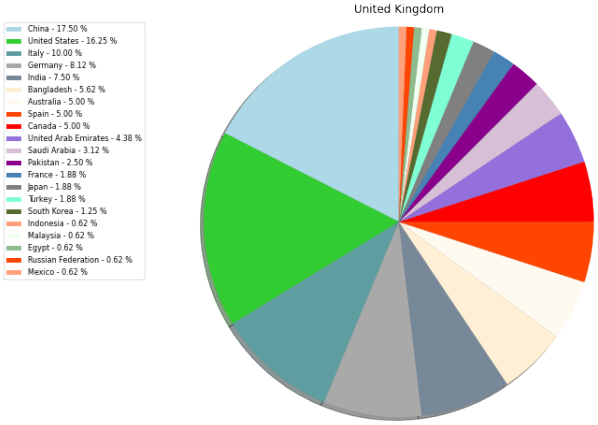
[37] Naveen Donthu, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, and Weng Marc Lim. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133:285–296, 2021. ISSN 0148-2963. doi: https://doi.org/10.1016/j.jbusres.2021.04.070. URL https://www.sciencedirect.com/science/article/pii/S0148296321003155.

A Collaborating Country Distribution

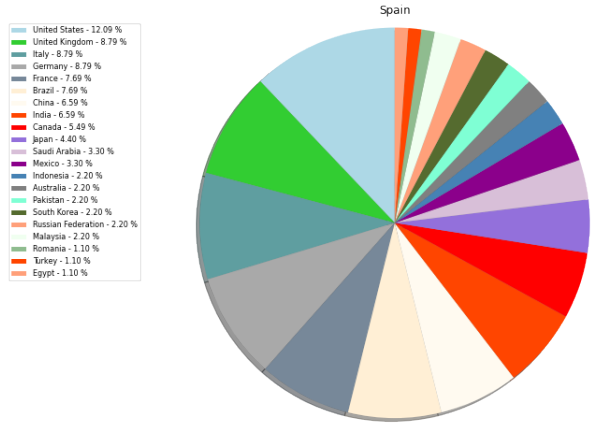


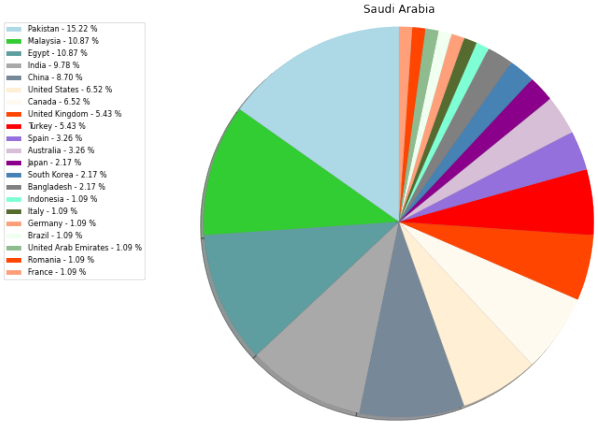




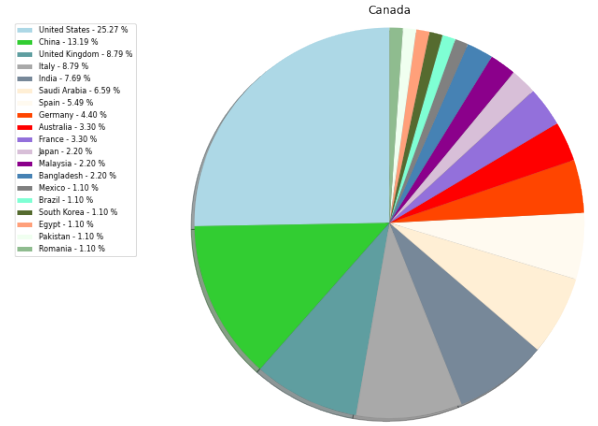


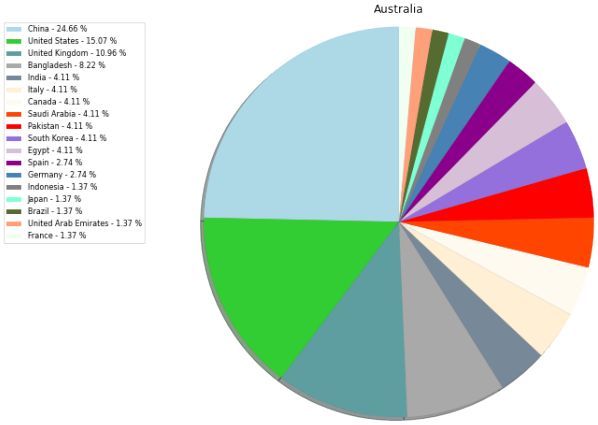
B Top 5 Keywords in Each Cluster



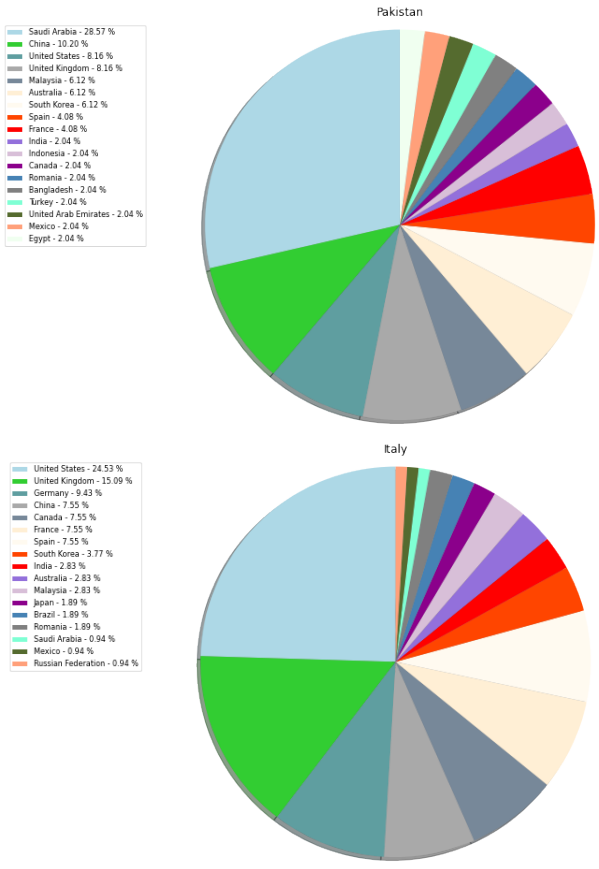


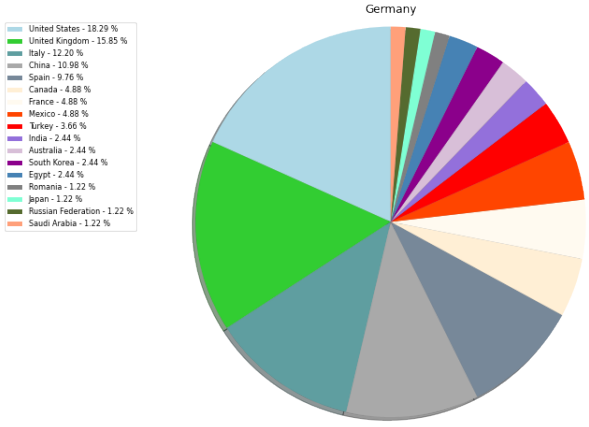
C Top 10 Most Significant Words Within Each Topic

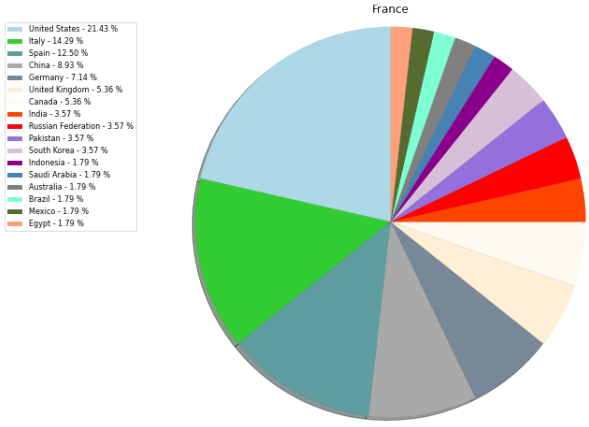


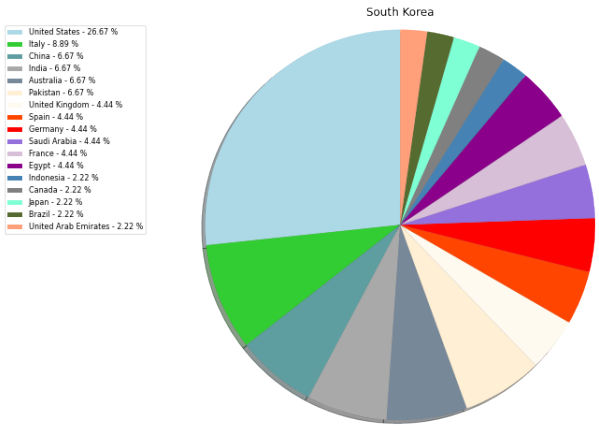


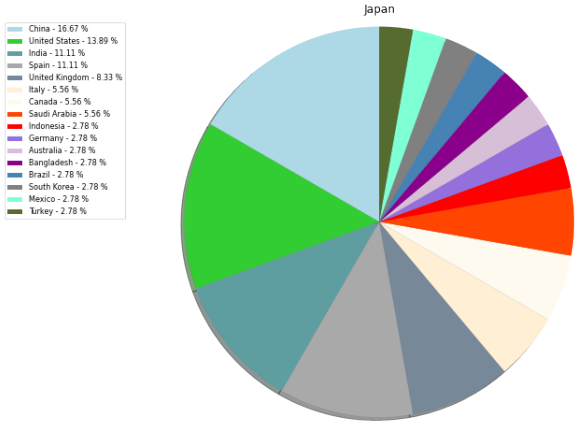
D Topics Distribution Within Each Country

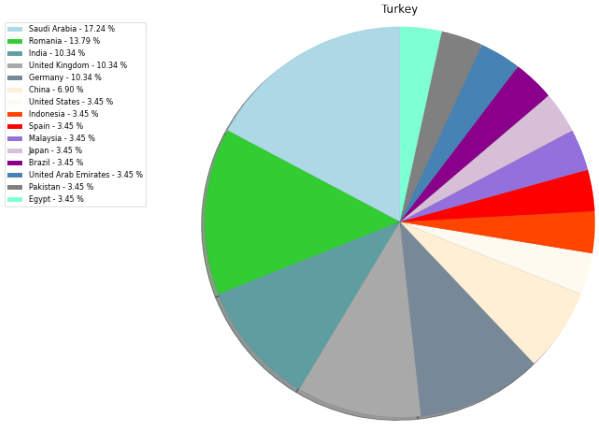

Pakistan
Saudi Arabia - 28.57 %
China - 10.20 %
United States - 8.16 %
United Kingdom - 8.16 %
Malaysia - 6.12 %
Australia - 6.12 %
South Korea - 6.12 %
Spain - 4.08 %
France - 4.08 %
India - 2.04 %
Indonesia - 2.04 %
Canada - 2.04 %
Romania - 2.04 %
Bangladesh - 2.04 %
Turkey - 2.04 %
United Arab Emirates - 2.04 %
Mexico - 2.04 %
Egypt - 2.04 %
Italy
United States - 24.53 %
United Kingdom - 15.09 %
Germany - 9.43 %
China - 7.55 %
Canada - 7.55 %
France - 7.55 %
Spain - 7.55 %
South Korea - 3.77 %
India - 2.83 %
Australia - 2.83 %
Malaysia - 2.83 %
Japan - 1.89 %
Brazil - 1.89 %
Romania - 1.89 %
Saudi Arabia - 0.94 %
Mexico - 0.94 %
Russian Federation - 0.94 %

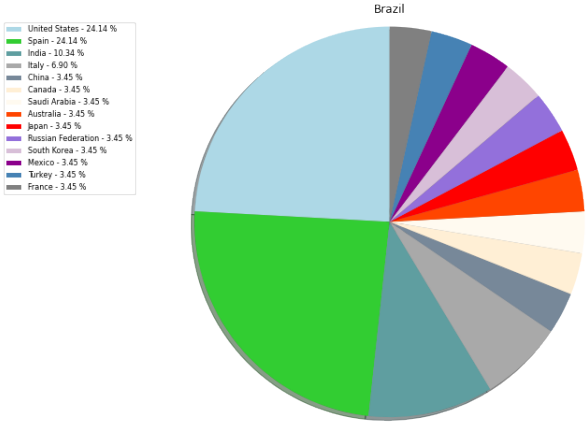

Germany
United States - 18.29 %
United Kingdom - 15.85 %
Italy - 12.20 %
China - 10.98 %
Spain - 9.76 %
Canada - 4.88 %
France - 4.88 %
Mexico - 4.88 %
Turkey - 3.66 %
India - 2.44 %
Australia - 2.44 %
South Korea - 2.44 %
Egypt - 2.44 %
Romania - 1.22 %
Japan - 1.22 %
Russian Federation - 1.22 %
Saudi Arabia - 1.22 %

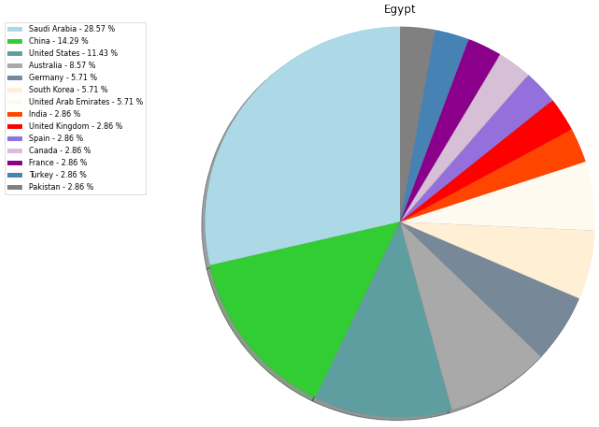

France
United States - 21.43 %
Italy - 14.29 %
Spain - 12.50 %
China - 8.93 %
Germany - 7.14 %
United Kingdom - 5.36 %
Canada - 5.36 %
India - 3.57 %
Russian Federation - 3.57 %
Pakistan - 3.57 %
South Korea - 3.57 %
Indonesia - 1.79 %
Saudi Arabia - 1.79 %
Australia - 1.79 %
Brazil - 1.79 %
Mexico - 1.79 %
Egypt - 1.79 %

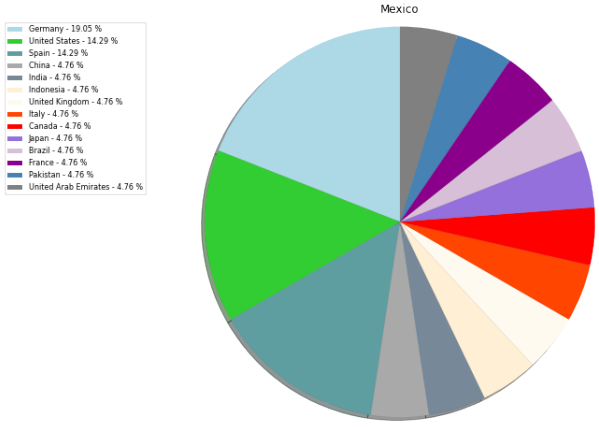

South Korea
United States - 26.67 %
Italy - 8.89 %
China - 6.67 %
India - 6.67 %
Australia - 6.67 %
Pakistan - 6.67 %
United Kingdom - 4.44 %
Spain - 4.44 %
Germany - 4.44 %
Saudi Arabia - 4.44 %
France - 4.44 %
Egypt - 4.44 %
Indonesia - 2.22 %
Canada - 2.22 %
Japan - 2.22 %
Brazil - 2.22 %
United Arab Emirates - 2.22 %

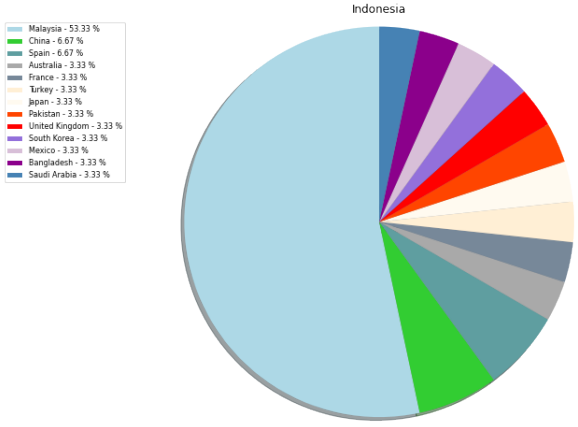

Japan
China - 16.67 %
United States - 13.89 %
India - 11.11 %
Spain - 11.11 %
United Kingdom - 8.33 %
Italy - 5.56 %
Canada - 5.56 %
Saudi Arabia - 5.56 %
Indonesia - 2.78 %
Germany - 2.78 %
Australia - 2.78 %
Bangladesh - 2.78 %
Brazil - 2.78 %
South Korea - 2.78 %
Mexico - 2.78 %
Turkey - 2.78 %

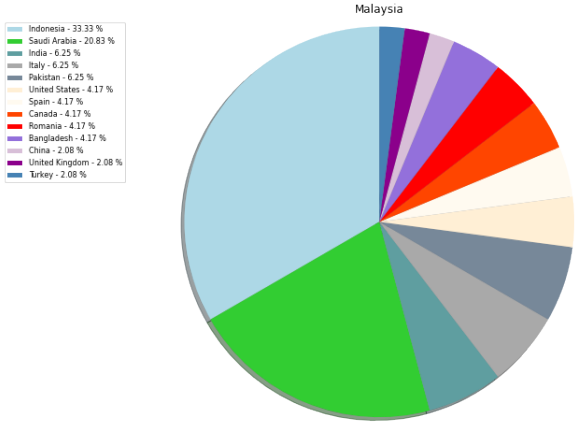

Turkey
Saudi Arabia - 17.24 %
Romania - 13.79 %
India - 10.34 %
United Kingdom - 10.34 %
Germany - 10.34 %
China - 6.90 %
United States - 3.45 %
Indonesia - 3.45 %
Spain - 3.45 %
Malaysia - 3.45 %
Japan - 3.45 %
Brazil - 3.45 %
United Arab Emirates - 3.45 %
Pakistan - 3.45 %
Egypt - 3.45 %

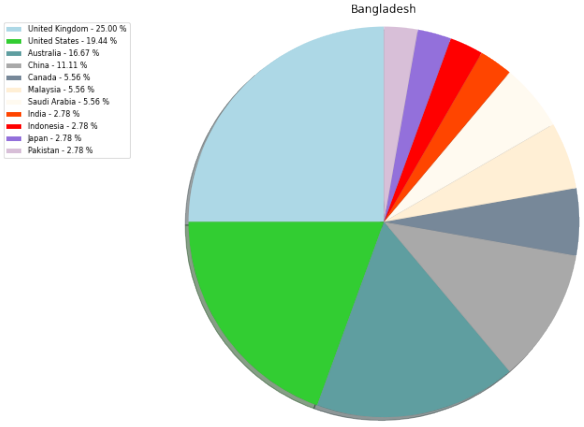

Brazil
United States - 24.14 %
Spain - 24.14 %
India - 10.34 %
Italy - 6.90 %
China - 3.45 %
Canada - 3.45 %
Saudi Arabia - 3.45 %
Australia - 3.45 %
Japan - 3.45 %
Russian Federation - 3.45 %
South Korea - 3.45 %
Mexico - 3.45 %
Turkey - 3.45 %
France - 3.45 %

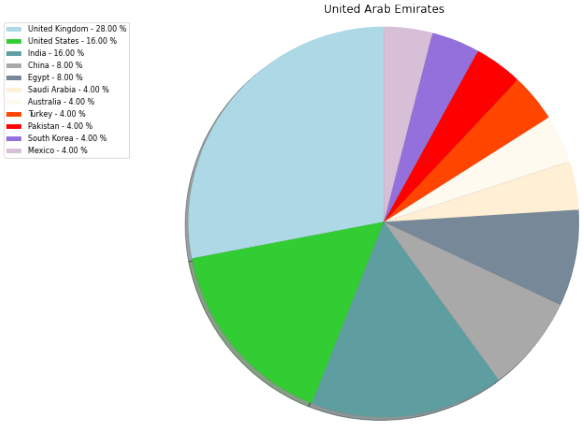

Egypt
Saudi Arabia - 28.57 %
China - 14.29 %
United States - 11.43 %
Australia - 8.57 %
Germany - 5.71 %
South Korea - 5.71 %
United Arab Emirates - 5.71 %
India - 2.86 %
United Kingdom - 2.86 %
Spain - 2.86 %
Canada - 2.86 %
France - 2.86 %
Turkey - 2.86 %
Pakistan - 2.86 %

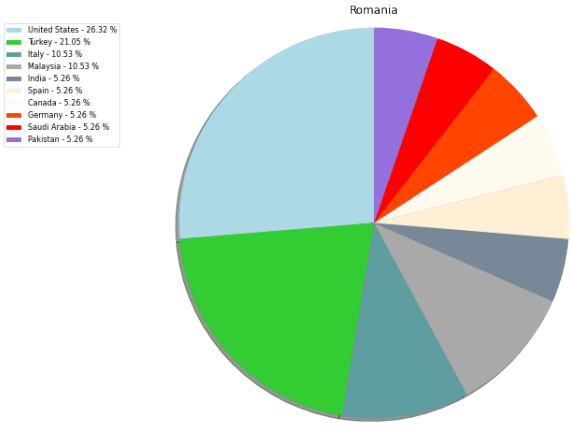

Mexico
Germany - 19.05 %
United States - 14.29 %
Spain - 14.29 %
China - 4.76 %
India - 4.76 %
Indonesia - 4.76 %
United Kingdom - 4.76 %
Italy - 4.76 %
Canada - 4.76 %
Japan - 4.76 %
Brazil - 4.76 %
France - 4.76 %
Pakistan - 4.76 %
United Arab Emirates - 4.76 %

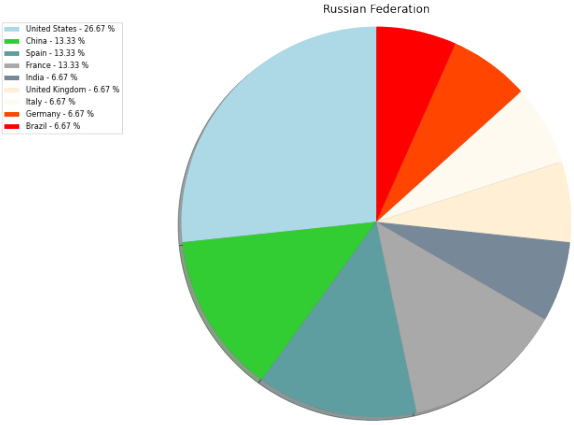

Indonesia
Malaysia - 53.33 %
China - 6.67 %
Spain - 6.67 %
Australia - 3.33 %
France - 3.33 %
Turkey - 3.33 %
Japan - 3.33 %
Pakistan - 3.33 %
United Kingdom - 3.33 %
South Korea - 3.33 %
Mexico - 3.33 %
Bangladesh - 3.33 %
Saudi Arabia - 3.33 %

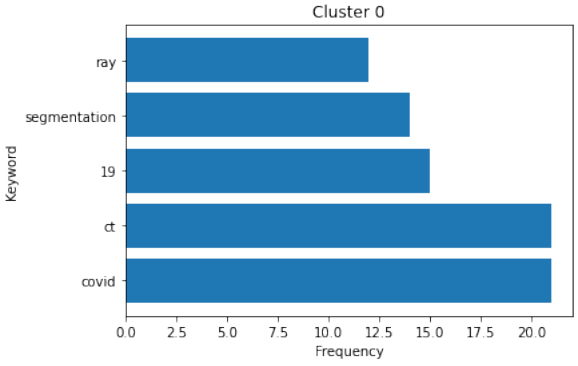

Malaysia
Indonesia - 33.33 %
Saudi Arabia - 20.83 %
India - 6.25 %
Italy - 6.25 %
Pakistan - 6.25 %
United States - 4.17 %
Spain - 4.17 %
Canada - 4.17 %
Romania - 4.17 %
Bangladesh - 4.17 %
China - 2.08 %
United Kingdom - 2.08 %
Turkey - 2.08 %

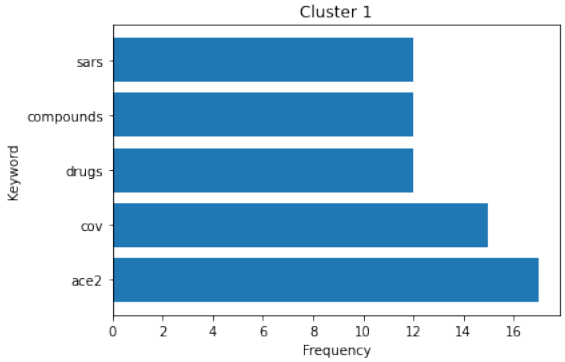

Bangladesh
United Kingdom - 25.00 %
United States - 19.44 %
Australia - 16.67 %
China - 11.11 %
Canada - 5.56 %
Malaysia - 5.56 %
Saudi Arabia - 5.56 %
India - 2.78 %
Indonesia - 2.78 %
Japan - 2.78 %
Pakistan - 2.78 %


United Arab Emirates
United Kingdom - 28.00 %
United States - 16.00 %
India - 16.00 %
China - 8.00 %
Egypt - 8.00 %
Saudi Arabia - 4.00 %
Australia - 4.00 %
Turkey - 4.00 %
Pakistan - 4.00 %
South Korea - 4.00 %
Mexico - 4.00 %

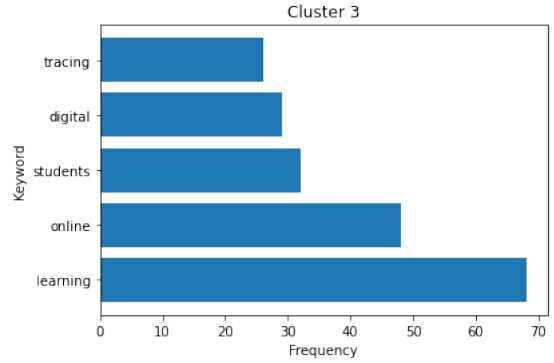

Romania
United States - 26.32 %
Turkey - 21.05 %
Italy - 10.53 %
Malaysia - 10.53 %
India - 5.26 %
Spain - 5.26 %
Canada - 5.26 %
Germany - 5.26 %
Saudi Arabia - 5.26 %
Pakistan - 5.26 %

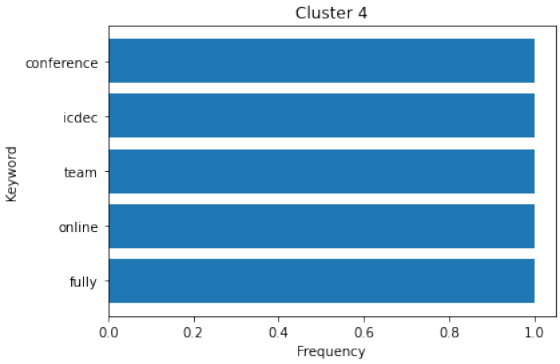

Russian Federation
United States - 26.67 %
China - 13.33 %
Spain - 13.33 %
France - 13.33 %
India - 6.67 %
United Kingdom - 6.67 %
Italy - 6.67 %
Germany - 6.67 %
Brazil - 6.67 %

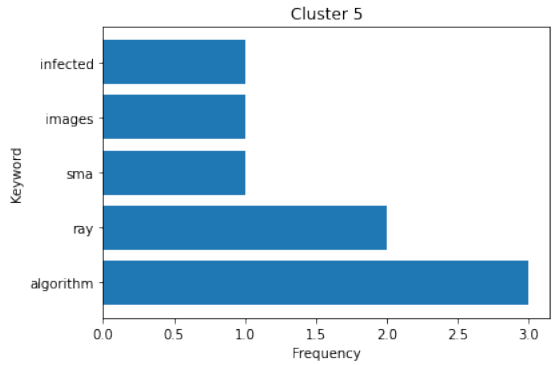

Cluster 0
ray
segmentation
19
ct
covid
Keyword
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Frequency

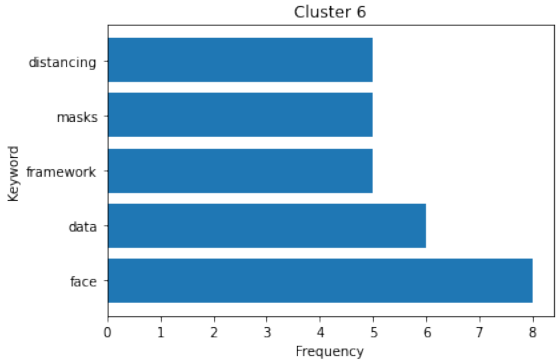

Cluster 1
sars
compounds
drugs
cov
ace2
Keyword
0
2
4
6
8
10
12
14
16
Frequency

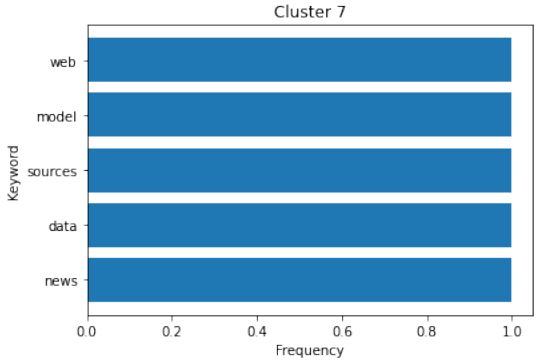

Cluster 2
cases
india
data
epidemic
model
Keyword
0
5
10
15
20
25
Frequency

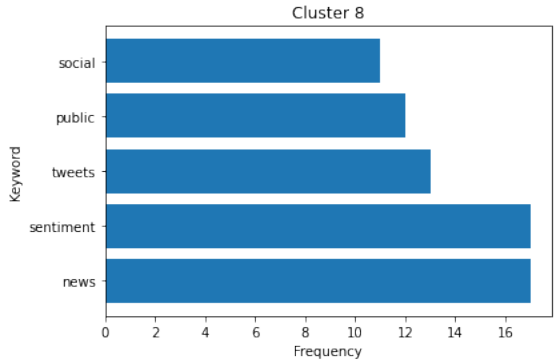

Cluster 3
tracing
digital
students
online
learning
Keyword
0
10
20
30
40
50
60
70
Frequency

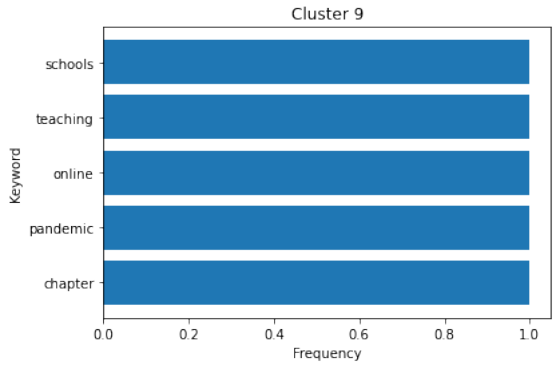

Cluster 4
conference
icdec
team
online
fully
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

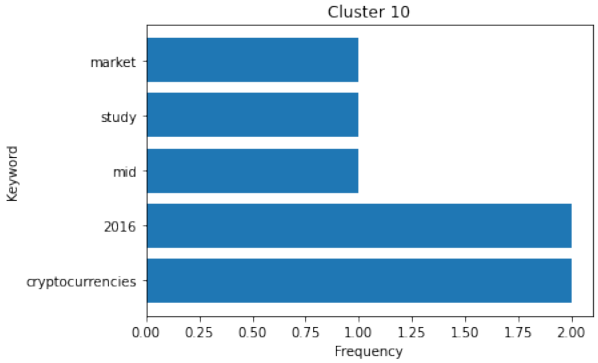

Cluster 5
infected
images
sma
ray
algorithm
Keyword
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Frequency

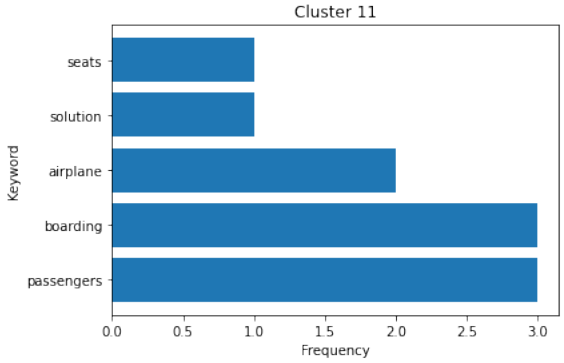

Cluster 6
distancing
masks
framework
data
face
Keyword
0
1
2
3
4
5
6
7
8
Frequency

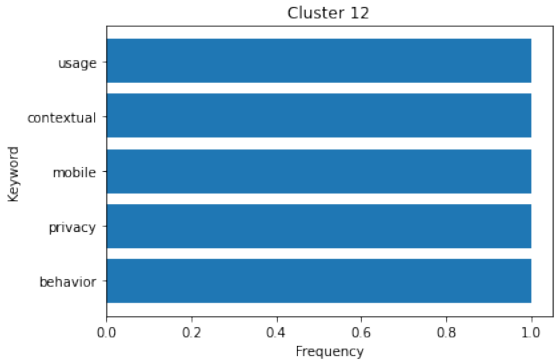

Cluster 7
web
model
sources
data
news
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

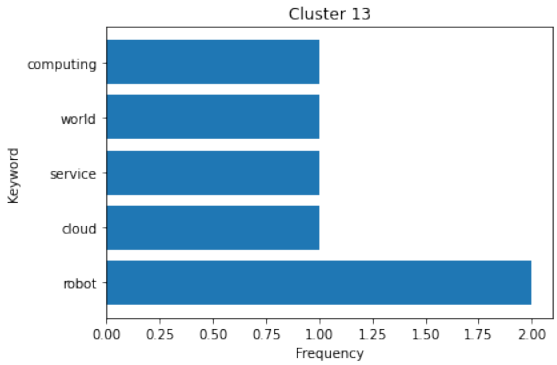

Cluster 8
social
public
tweets
sentiment
news
Keyword
0
2
4
6
8
10
12
14
16
Frequency

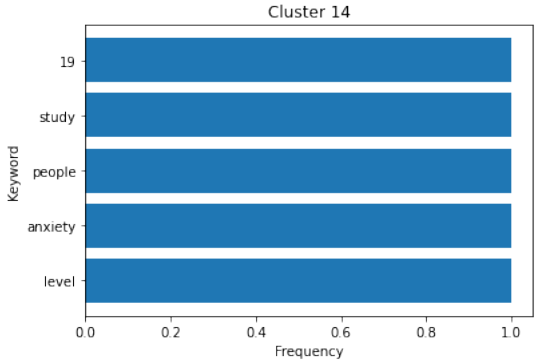

Cluster 9
schools
teaching
online
pandemic
chapter
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

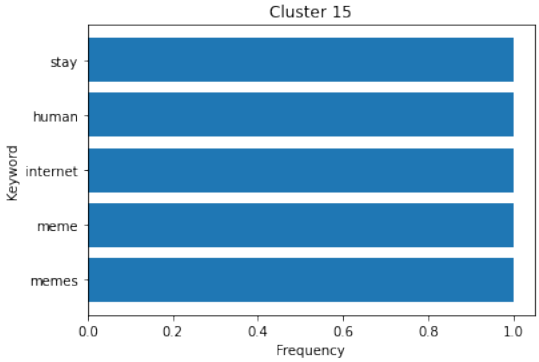

Cluster 10
market
study
mid
2016
cryptocurrencies
Keyword
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00
Frequency

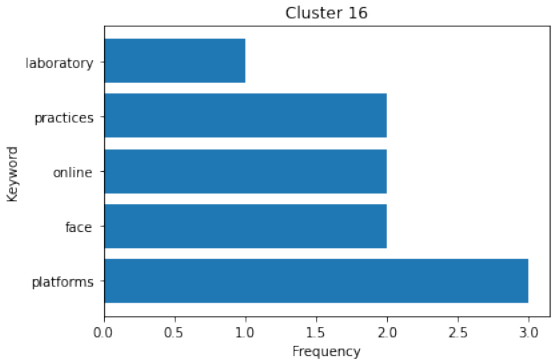

Cluster 11
seats
solution
airplane
boarding
passengers
Keyword
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Frequency

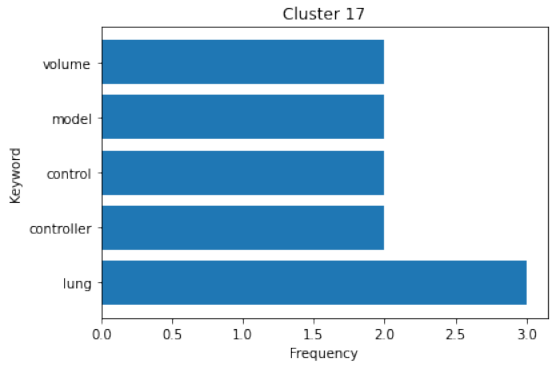

Cluster 12
usage
contextual
mobile
privacy
behavior
Keyword
0.0 0.2 0.4 0.6 0.8 1.0
Frequency

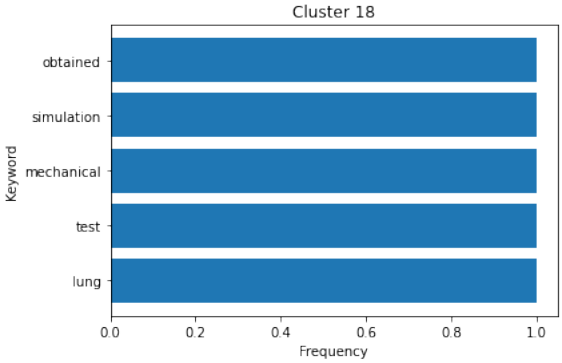

Cluster 13
computing
world
service
cloud
robot
Keyword
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00
Frequency

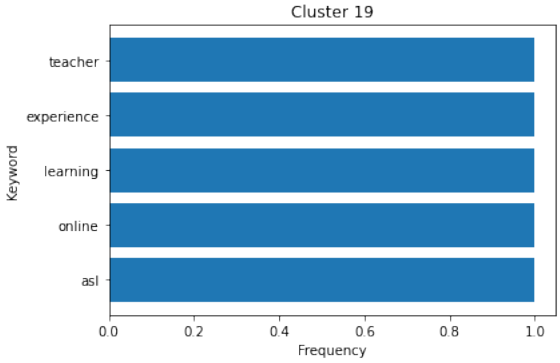

Cluster 14
19
study
people
anxiety
level
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

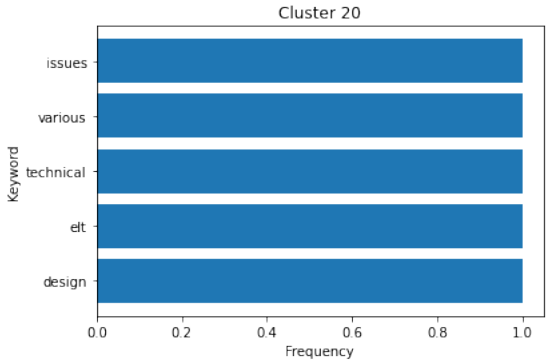

Cluster 15
stay
human
internet
meme
memes
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

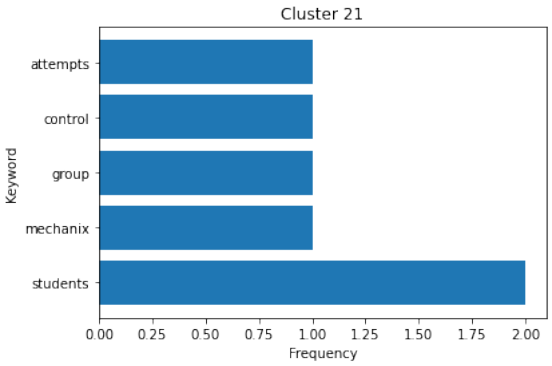

Cluster 16
laboratory
practices
online
face
platforms
Keyword
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Frequency

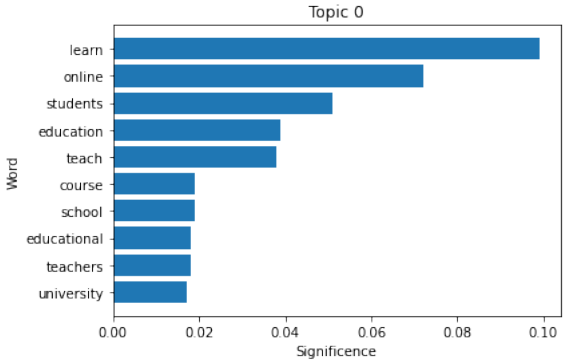

Cluster 17
volume
model
control
controller
lung
Keyword
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Frequency

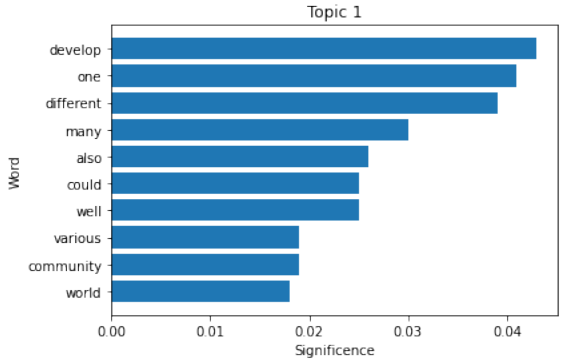

Cluster 18
obtained
simulation
mechanical
test
lung
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

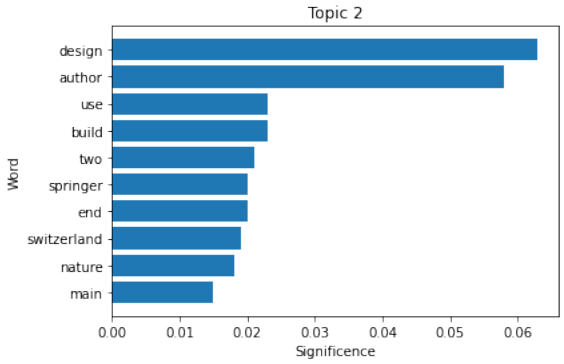

Cluster 19
teacher
experience
learning
online
asl
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

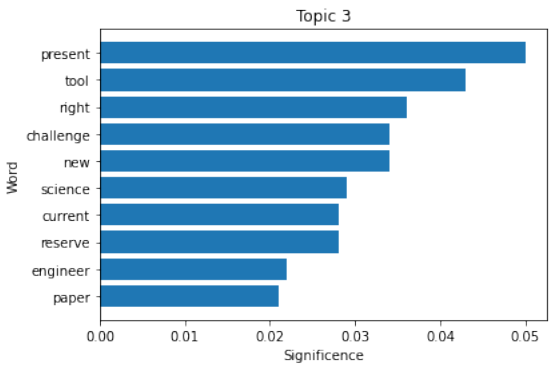

Cluster 20
issues
various
technical
elt
design
Keyword
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

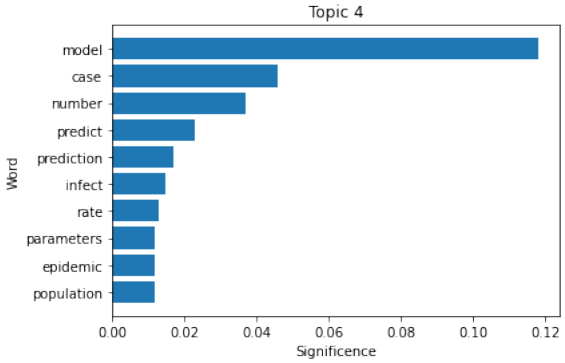

Cluster 21
attempts
control
group
mechanix
students
Keyword
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Frequency

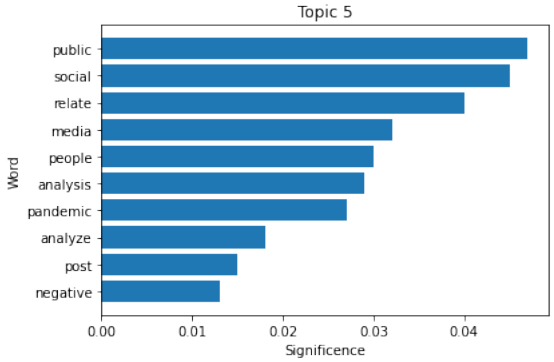

Topic 0
learn
online
students
education
teach
course
school
educational
teachers
university
Word
0.00
0.02
0.04
0.06
0.08
0.10
Significence

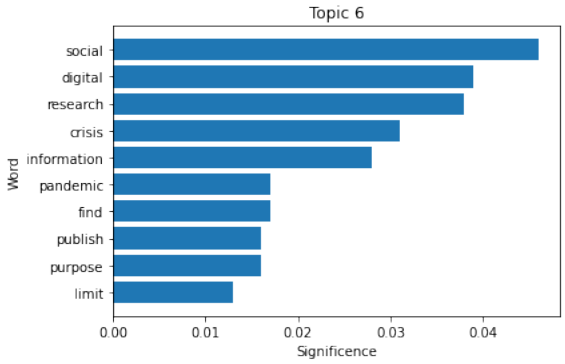

Topic 1
develop
one
different
many
also
could
well
various
community
world
Word
0.00
0.01
0.02
0.03
0.04
Significence

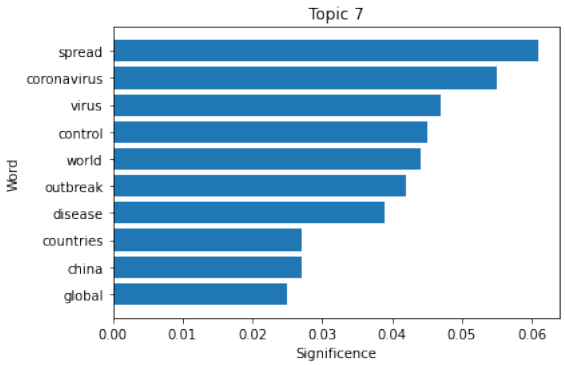

Topic 2
design
author
use
build
two
springer
end
switzerland
nature
main
Word
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Significence

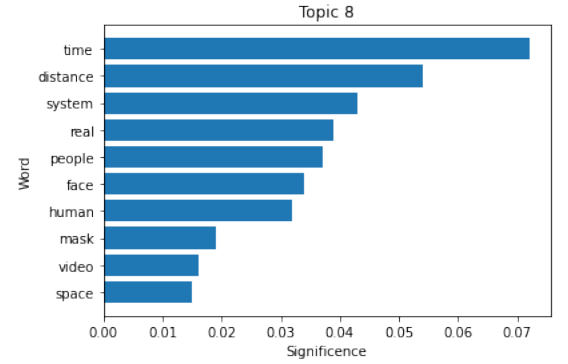

Topic 3
present
tool
right
challenge
new
science
current
reserve
engineer
paper
Word
0.00
0.01
0.02
0.03
0.04
0.05
Significence

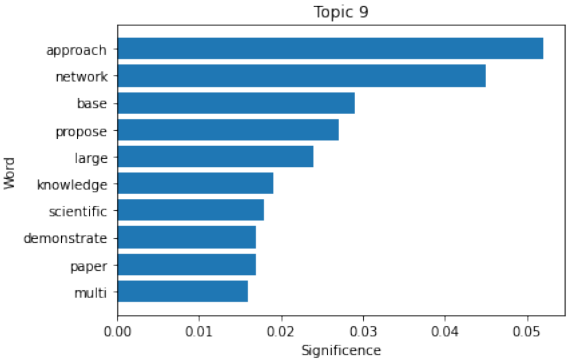

Topic 4
model
case
number
predict
prediction
infect
rate
parameters
epidemic
population
Word
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Significence

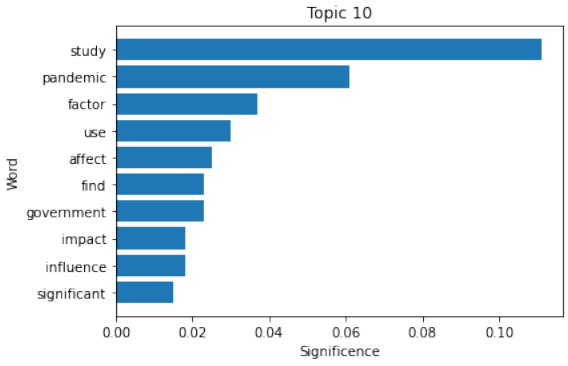

Topic 5
public
social
relate
media
people
analysis
pandemic
analyze
post
negative
Word
0.00
0.01
0.02
0.03
0.04
Significence

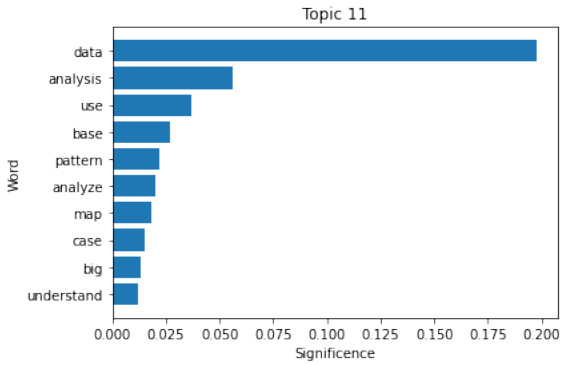

Topic 6
social
digital
research
crisis
information
pandemic
find
publish
purpose
limit
Word
0.00
0.01
0.02
0.03
0.04
Significence

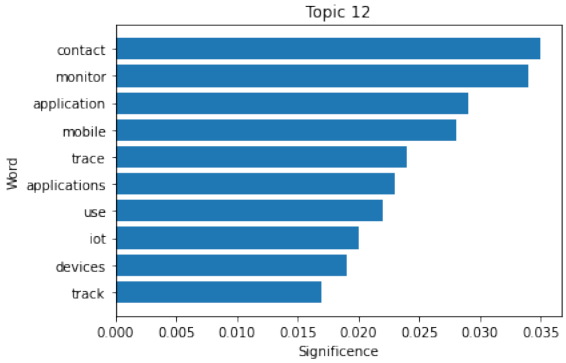

Topic 7
spread
coronavirus
virus
control
world
outbreak
disease
countries
china
global
Word
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Significence

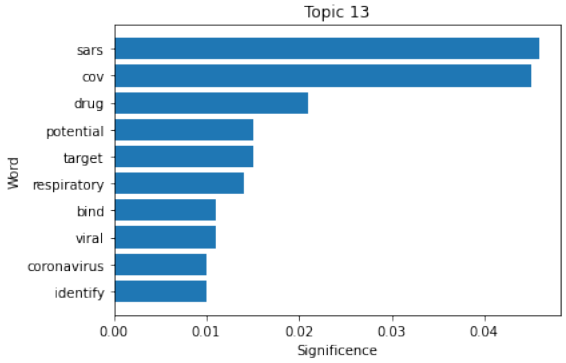

Topic 8
time
distance
system
real
people
face
human
mask
video
space
Word
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Significence

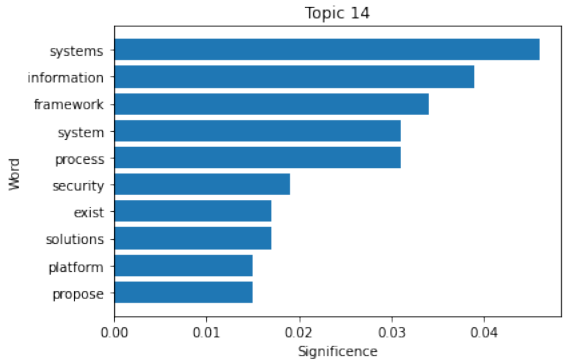

Topic 9
approach
network
base
propose
large
knowledge
scientific
demonstrate
paper
multi
Word
0.00
0.01
0.02
0.03
0.04
0.05
Significence

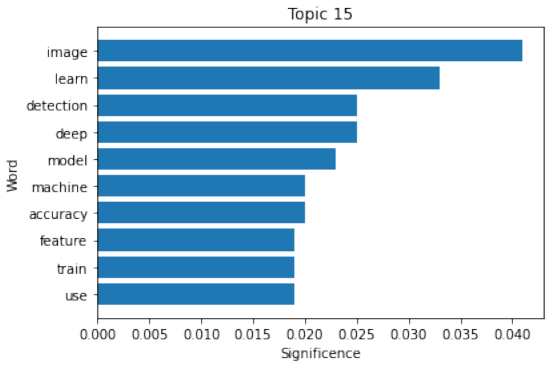

Topic 10
study
pandemic
factor
use
affect
find
government
impact
influence
significant
Word
0.00
0.02
0.04
0.06
0.08
0.10
Significence

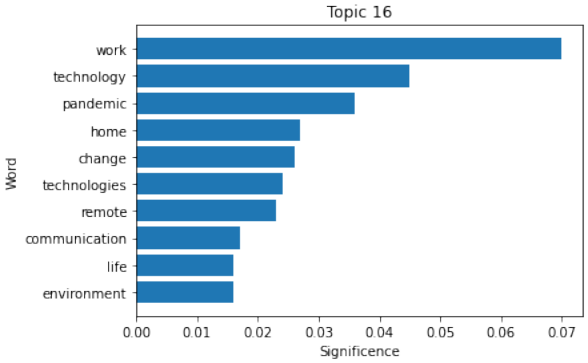

Topic 11
data
analysis
use
base
pattern
analyze
map
case
big
understand
Word
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
Significence

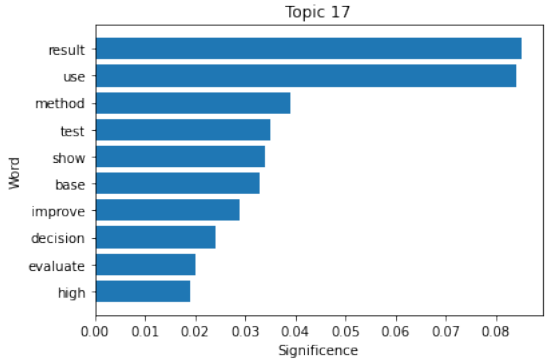

Topic 12
contact
monitor
application
mobile
trace
applications
use
iot
devices
track
Word
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
Significence

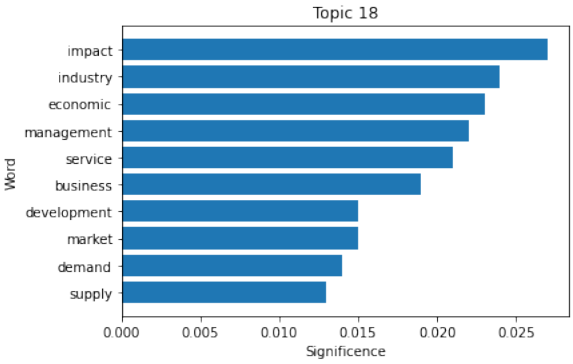

Topic 13
sars
cov
drug
potential
target
respiratory
bind
viral
coronavirus
identify
Word
0.00
0.01
0.02
0.03
0.04
Significence

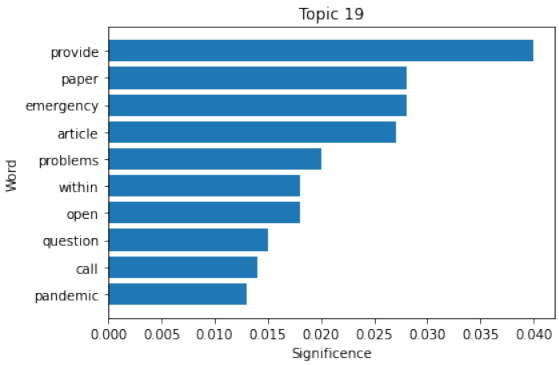

Topic 14
systems
information
framework
system
process
security
exist
solutions
platform
propose
Word
0.00
0.01
0.02
0.03
0.04
Significence

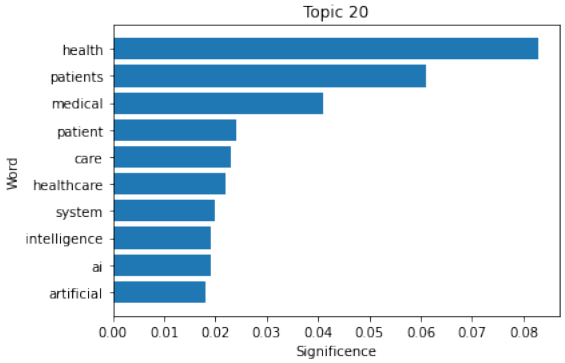

Topic 15
image
learn
detection
deep
model
machine
accuracy
feature
train
use
Word
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
Significence

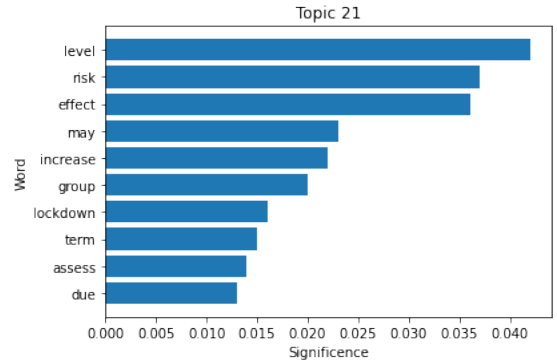

Topic 16
work
technology
pandemic
home
change
technologies
remote
communication
life
environment
Word
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Significence

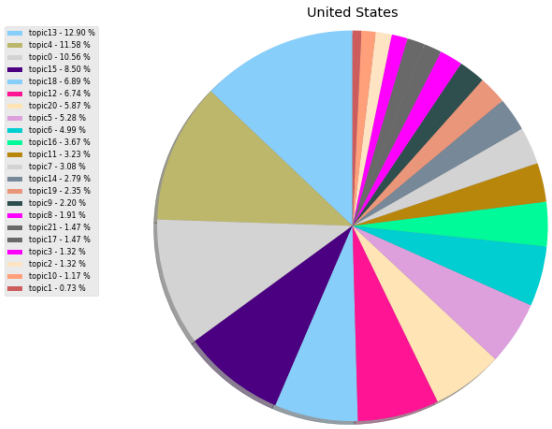

Topic 17
result
use
method
test
show
base
improve
decision
evaluate
high
Word
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Significence

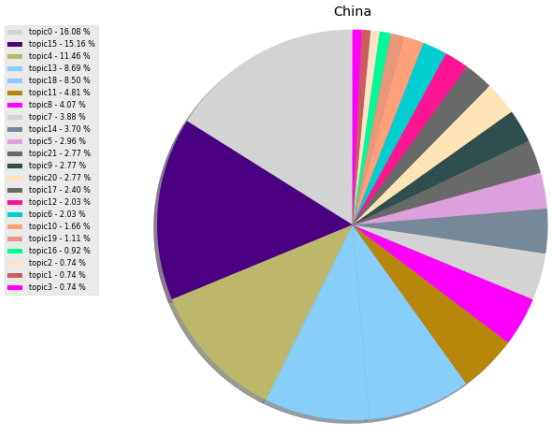

Topic 18
impact
industry
economic
management
service
business
development
market
demand
supply
Word
0.000
0.005
0.010
0.015
0.020
0.025
Significence

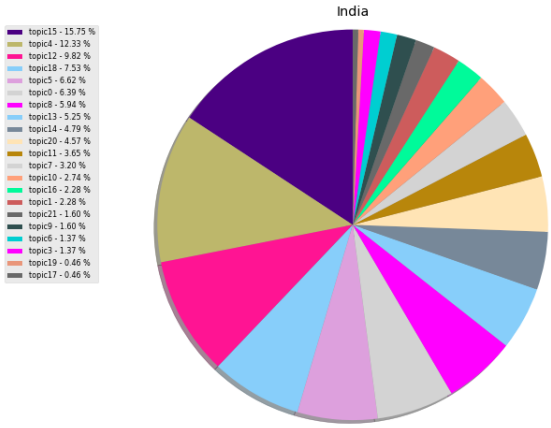

Topic 19
provide
paper
emergency
article
problems
within
open
question
call
pandemic
Word
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
Significence

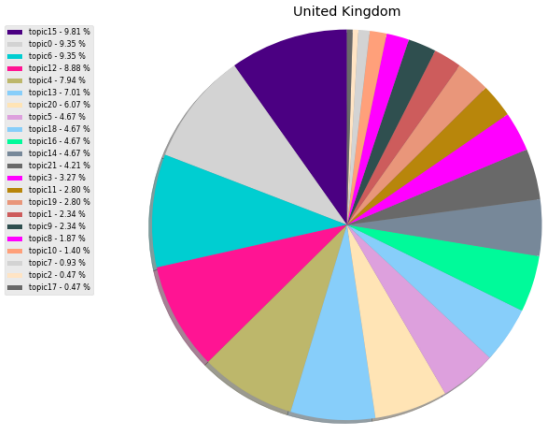

Topic 20
health
patients
medical
patient
care
healthcare
system
intelligence
ai
artificial
Word
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08
Significence

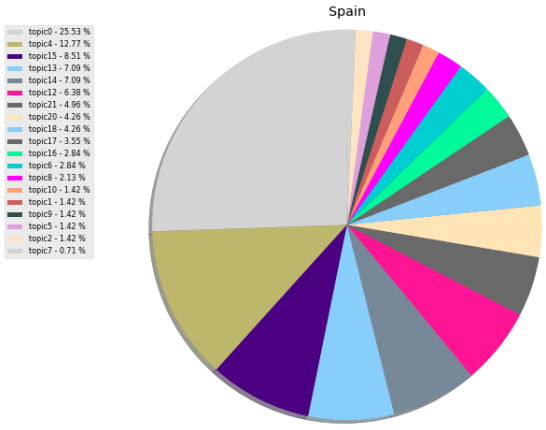

Topic 21
level
risk
effect
may
increase
group
lockdown
term
assess
due
Word
0.000 0.005 0.010 0.015 0.020 0.025 0.030 0.035 0.040
Significence

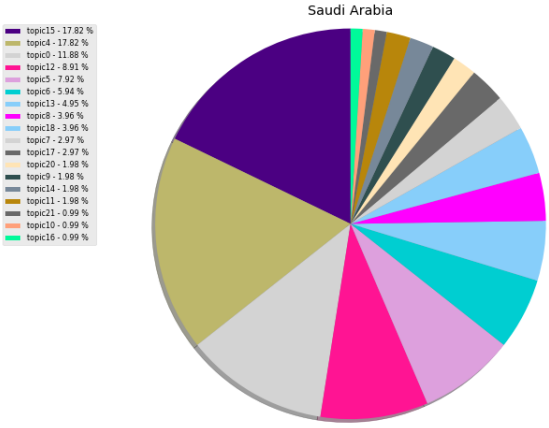

United States
topic13 - 12.90 %
topic4 - 11.58 %
topic0 - 10.56 %
topic15 - 8.50 %
topic18 - 6.89 %
topic12 - 6.74 %
topic20 - 5.87 %
topic5 - 5.28 %
topic6 - 4.99 %
topic16 - 3.67 %
topic11 - 3.23 %
topic7 - 3.08 %
topic14 - 2.79 %
topic19 - 2.35 %
topic9 - 2.20 %
topic8 - 1.91 %
topic21 - 1.47 %
topic17 - 1.47 %
topic3 - 1.32 %
topic2 - 1.32 %
topic10 - 1.17 %
topic1 - 0.73 %

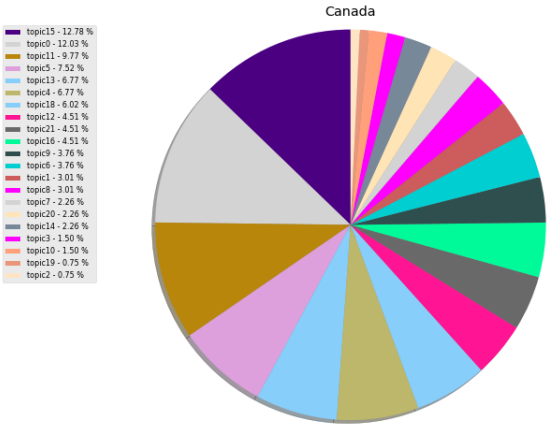

China
topic0 - 16.08 %
topic15 - 15.16 %
topic4 - 11.46 %
topic13 - 8.69 %
topic18 - 8.50 %
topic11 - 4.81 %
topic8 - 4.07 %
topic7 - 3.88 %
topic14 - 3.70 %
topic5 - 2.96 %
topic21 - 2.77 %
topic9 - 2.77 %
topic20 - 2.77 %
topic17 - 2.40 %
topic12 - 2.03 %
topic6 - 2.03 %
topic10 - 1.66 %
topic19 - 1.11 %
topic16 - 0.92 %
topic2 - 0.74 %
topic1 - 0.74 %
topic3 - 0.74 %

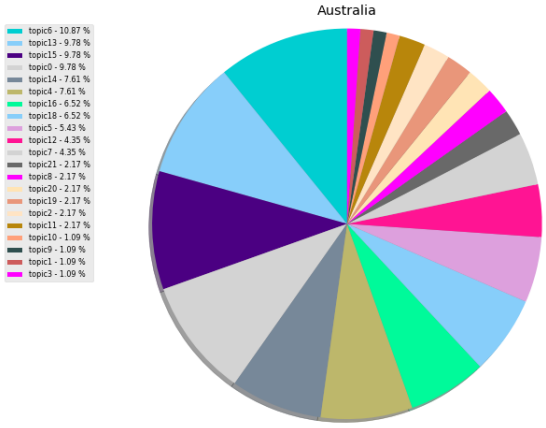

India
topic15 - 15.75 %
topic4 - 12.33 %
topic12 - 9.82 %
topic18 - 7.53 %
topic5 - 6.62 %
topic0 - 6.39 %
topic8 - 5.94 %
topic13 - 5.25 %
topic14 - 4.79 %
topic20 - 4.57 %
topic11 - 3.65 %
topic7 - 3.20 %
topic10 - 2.74 %
topic16 - 2.28 %
topic1 - 2.28 %
topic21 - 1.60 %
topic9 - 1.60 %
topic6 - 1.37 %
topic3 - 1.37 %
topic19 - 0.46 %
topic17 - 0.46 %

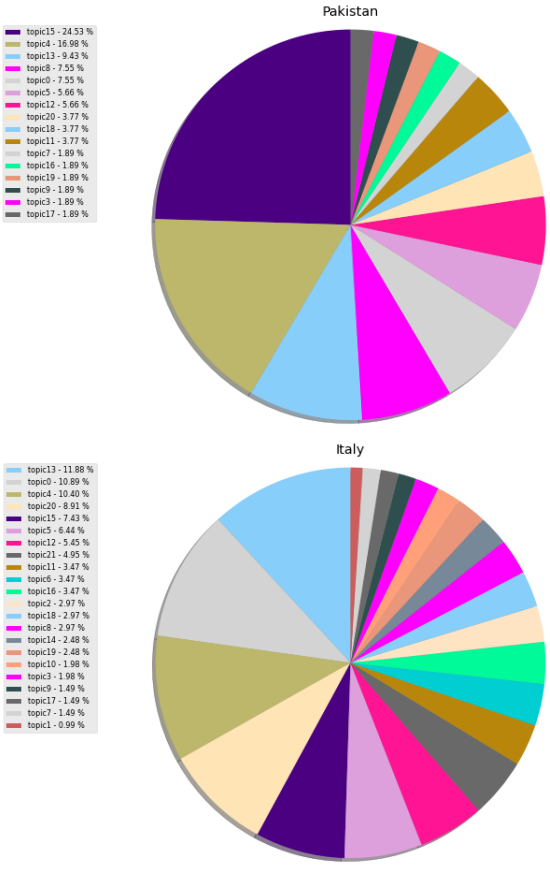

United Kingdom
topic15 - 9.81 %
topic0 - 9.35 %
topic6 - 9.35 %
topic12 - 8.88 %
topic4 - 7.94 %
topic13 - 7.01 %
topic20 - 6.07 %
topic5 - 4.67 %
topic18 - 4.67 %
topic16 - 4.67 %
topic14 - 4.67 %
topic21 - 4.21 %
topic3 - 3.27 %
topic11 - 2.80 %
topic19 - 2.80 %
topic1 - 2.34 %
topic9 - 2.34 %
topic8 - 1.87 %
topic10 - 1.40 %
topic7 - 0.93 %
topic2 - 0.47 %
topic17 - 0.47 %

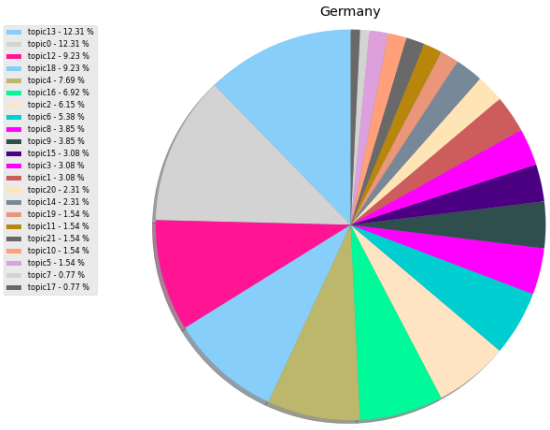

Spain
topic0 - 25.53 %
topic4 - 12.77 %
topic15 - 8.51 %
topic13 - 7.09 %
topic14 - 7.09 %
topic12 - 6.38 %
topic21 - 4.96 %
topic20 - 4.26 %
topic18 - 4.26 %
topic17 - 3.55 %
topic16 - 2.84 %
topic6 - 2.84 %
topic8 - 2.13 %
topic10 - 1.42 %
topic1 - 1.42 %
topic9 - 1.42 %
topic5 - 1.42 %
topic2 - 1.42 %
topic7 - 0.71 %

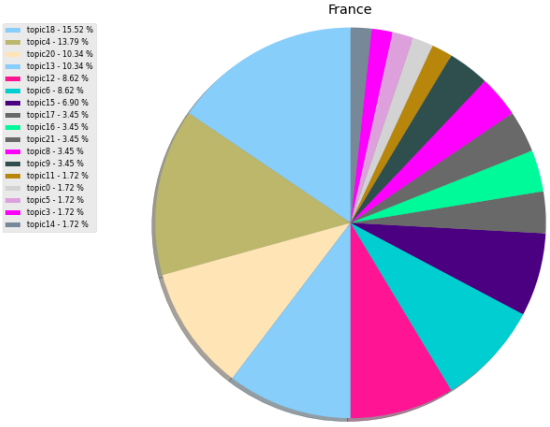

Saudi Arabia
topic15 - 17.82 %
topic4 - 17.82 %
topic0 - 11.88 %
topic12 - 8.91 %
topic5 - 7.92 %
topic6 - 5.94 %
topic13 - 4.95 %
topic8 - 3.96 %
topic18 - 3.96 %
topic7 - 2.97 %
topic17 - 2.97 %
topic20 - 1.98 %
topic9 - 1.98 %
topic14 - 1.98 %
topic11 - 1.98 %
topic21 - 0.99 %
topic10 - 0.99 %
topic16 - 0.99 %

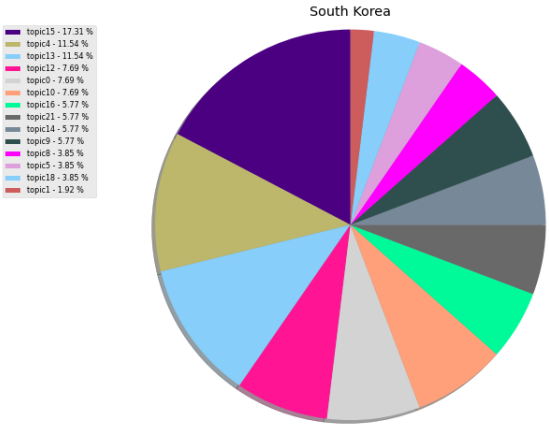

Canada
topic15 - 12.78 %
topic0 - 12.03 %
topic11 - 9.77 %
topic5 - 7.52 %
topic13 - 6.77 %
topic4 - 6.77 %
topic18 - 6.02 %
topic12 - 4.51 %
topic21 - 4.51 %
topic16 - 4.51 %
topic9 - 3.76 %
topic6 - 3.76 %
topic1 - 3.01 %
topic8 - 3.01 %
topic7 - 2.26 %
topic20 - 2.26 %
topic14 - 2.26 %
topic3 - 1.50 %
topic10 - 1.50 %
topic19 - 0.75 %
topic2 - 0.75 %

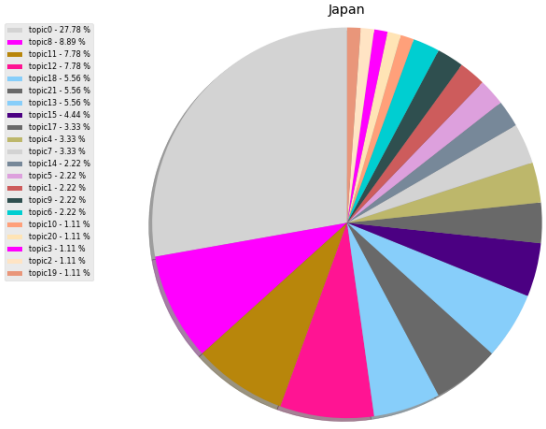

Australia
topic6 - 10.87 %
topic13 - 9.78 %
topic15 - 9.78 %
topic0 - 9.78 %
topic14 - 7.61 %
topic4 - 7.61 %
topic16 - 6.52 %
topic18 - 6.52 %
topic5 - 5.43 %
topic12 - 4.35 %
topic7 - 4.35 %
topic21 - 2.17 %
topic8 - 2.17 %
topic20 - 2.17 %
topic19 - 2.17 %
topic2 - 2.17 %
topic11 - 2.17 %
topic10 - 1.09 %
topic9 - 1.09 %
topic1 - 1.09 %
topic3 - 1.09 %

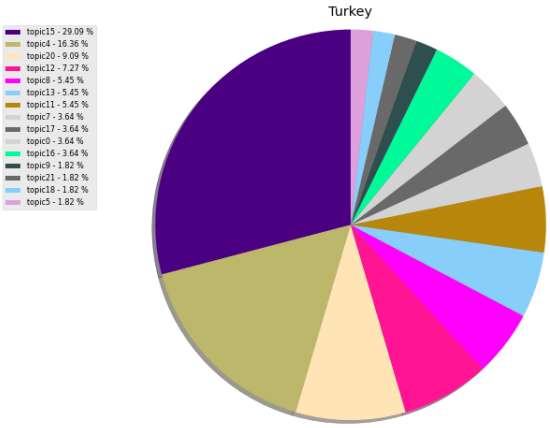

Pakistan
topic15 - 24.53 %
topic4 - 16.98 %
topic13 - 9.43 %
topic8 - 7.55 %
topic0 - 7.55 %
topic5 - 5.66 %
topic12 - 5.66 %
topic20 - 3.77 %
topic18 - 3.77 %
topic11 - 3.77 %
topic7 - 1.89 %
topic16 - 1.89 %
topic19 - 1.89 %
topic9 - 1.89 %
topic3 - 1.89 %
topic17 - 1.89 %
Italy
topic13 - 11.88 %
topic0 - 10.89 %
topic4 - 10.40 %
topic20 - 8.91 %
topic15 - 7.43 %
topic5 - 6.44 %
topic12 - 5.45 %
topic21 - 4.95 %
topic11 - 3.47 %
topic6 - 3.47 %
topic16 - 3.47 %
topic2 - 2.97 %
topic18 - 2.97 %
topic8 - 2.97 %
topic14 - 2.48 %
topic19 - 2.48 %
topic10 - 1.98 %
topic3 - 1.98 %
topic9 - 1.49 %
topic17 - 1.49 %
topic7 - 1.49 %
topic1 - 0.99 %

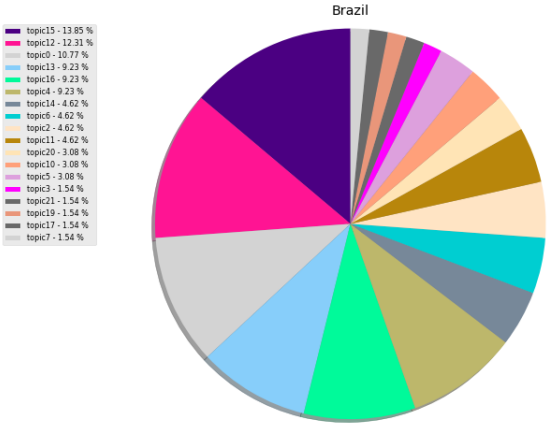

Germany
topic13 - 12.31 %
topic0 - 12.31 %
topic12 - 9.23 %
topic18 - 9.23 %
topic4 - 7.69 %
topic16 - 6.92 %
topic2 - 6.15 %
topic6 - 5.38 %
topic8 - 3.85 %
topic9 - 3.85 %
topic15 - 3.08 %
topic3 - 3.08 %
topic1 - 3.08 %
topic20 - 2.31 %
topic14 - 2.31 %
topic19 - 1.54 %
topic11 - 1.54 %
topic21 - 1.54 %
topic10 - 1.54 %
topic5 - 1.54 %
topic7 - 0.77 %
topic17 - 0.77 %

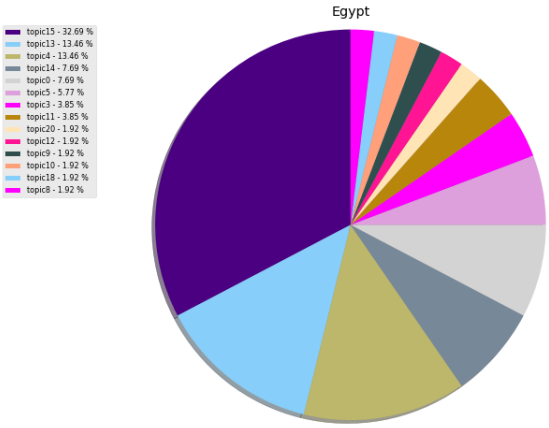

France
topic18 - 15.52 %
topic4 - 13.79 %
topic20 - 10.34 %
topic13 - 10.34 %
topic12 - 8.62 %
topic6 - 8.62 %
topic15 - 6.90 %
topic17 - 3.45 %
topic16 - 3.45 %
topic21 - 3.45 %
topic8 - 3.45 %
topic9 - 3.45 %
topic11 - 1.72 %
topic0 - 1.72 %
topic5 - 1.72 %
topic3 - 1.72 %
topic14 - 1.72 %

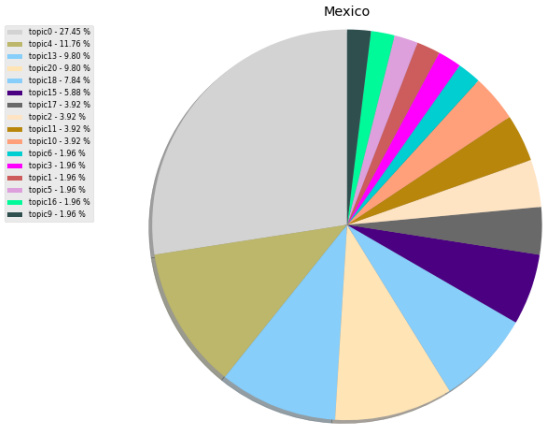

South Korea
topic15 - 17.31 %
topic4 - 11.54 %
topic13 - 11.54 %
topic12 - 7.69 %
topic0 - 7.69 %
topic10 - 7.69 %
topic16 - 5.77 %
topic21 - 5.77 %
topic14 - 5.77 %
topic9 - 5.77 %
topic8 - 3.85 %
topic5 - 3.85 %
topic18 - 3.85 %
topic1 - 1.92 %

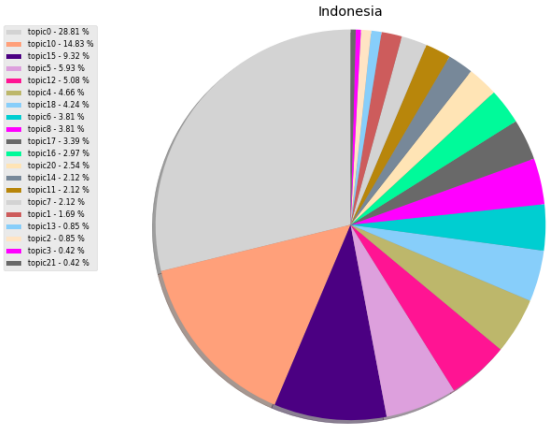

Japan
topic0 - 27.78 %
topic8 - 8.89 %
topic11 - 7.78 %
topic12 - 7.78 %
topic18 - 5.56 %
topic21 - 5.56 %
topic13 - 5.56 %
topic15 - 4.44 %
topic17 - 3.33 %
topic4 - 3.33 %
topic7 - 3.33 %
topic14 - 2.22 %
topic5 - 2.22 %
topic1 - 2.22 %
topic9 - 2.22 %
topic6 - 2.22 %
topic10 - 1.11 %
topic20 - 1.11 %
topic3 - 1.11 %
topic2 - 1.11 %
topic19 - 1.11 %

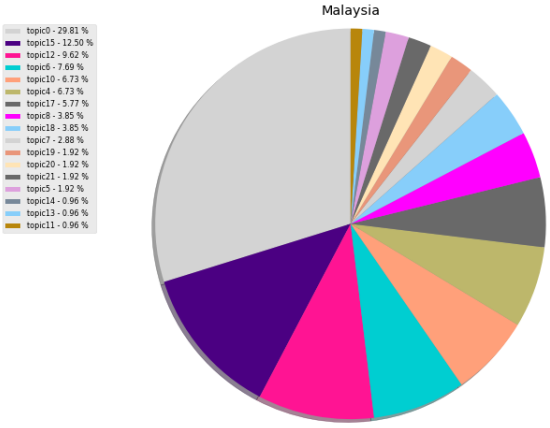

Turkey
topic15 - 29.09 %
topic4 - 16.36 %
topic20 - 9.09 %
topic12 - 7.27 %
topic8 - 5.45 %
topic13 - 5.45 %
topic11 - 5.45 %
topic7 - 3.64 %
topic17 - 3.64 %
topic0 - 3.64 %
topic16 - 3.64 %
topic9 - 1.82 %
topic21 - 1.82 %
topic18 - 1.82 %
topic5 - 1.82 %

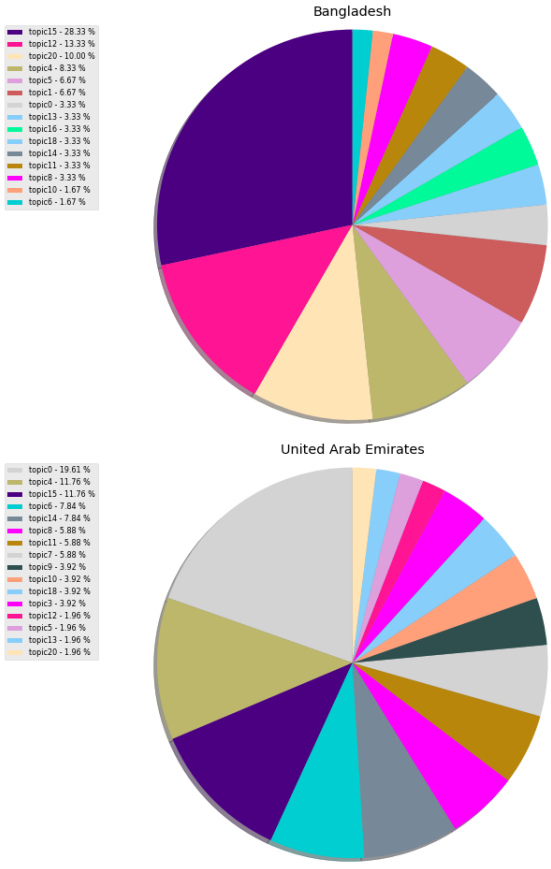

Brazil
topic15 - 13.85 %
topic12 - 12.31 %
topic0 - 10.77 %
topic13 - 9.23 %
topic16 - 9.23 %
topic4 - 9.23 %
topic14 - 4.62 %
topic6 - 4.62 %
topic2 - 4.62 %
topic11 - 4.62 %
topic20 - 3.08 %
topic10 - 3.08 %
topic5 - 3.08 %
topic3 - 1.54 %
topic21 - 1.54 %
topic19 - 1.54 %
topic17 - 1.54 %
topic7 - 1.54 %

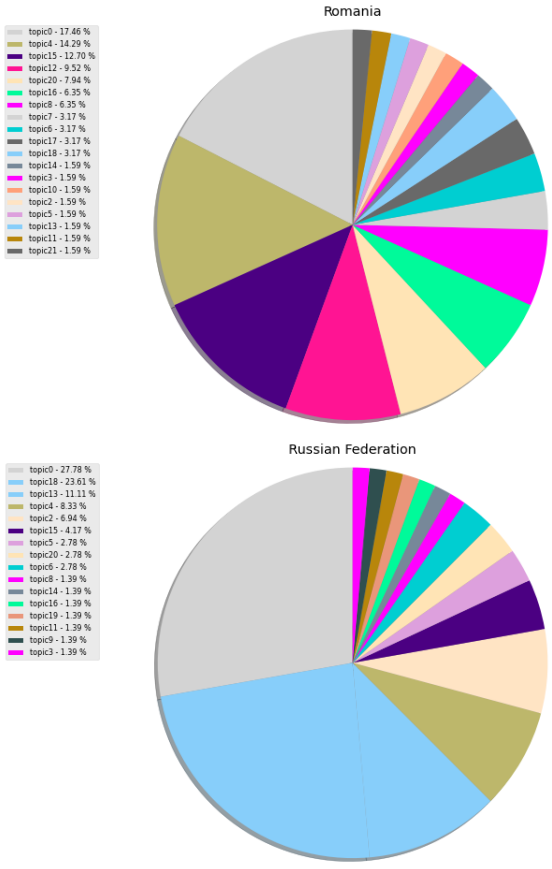

Egypt
topic15 - 32.69 %
topic13 - 13.46 %
topic4 - 13.46 %
topic14 - 7.69 %
topic0 - 7.69 %
topic5 - 5.77 %
topic3 - 3.85 %
topic11 - 3.85 %
topic20 - 1.92 %
topic12 - 1.92 %
topic9 - 1.92 %
topic10 - 1.92 %
topic18 - 1.92 %
topic8 - 1.92 %


Mexico
topic0 - 27.45 %
topic4 - 11.76 %
topic13 - 9.80 %
topic20 - 9.80 %
topic18 - 7.84 %
topic15 - 5.88 %
topic17 - 3.92 %
topic2 - 3.92 %
topic11 - 3.92 %
topic10 - 3.92 %
topic6 - 1.96 %
topic3 - 1.96 %
topic1 - 1.96 %
topic5 - 1.96 %
topic16 - 1.96 %
topic9 - 1.96 %


Indonesia
topic0 - 28.81 %
topic10 - 14.83 %
topic15 - 9.32 %
topic5 - 5.93 %
topic12 - 5.08 %
topic4 - 4.66 %
topic18 - 4.24 %
topic6 - 3.81 %
topic8 - 3.81 %
topic17 - 3.39 %
topic16 - 2.97 %
topic20 - 2.54 %
topic14 - 2.12 %
topic11 - 2.12 %
topic7 - 2.12 %
topic1 - 1.69 %
topic13 - 0.85 %
topic2 - 0.85 %
topic3 - 0.42 %
topic21 - 0.42 %


Malaysia
topic0 - 29.81 %
topic15 - 12.50 %
topic12 - 9.62 %
topic6 - 7.69 %
topic10 - 6.73 %
topic4 - 6.73 %
topic17 - 5.77 %
topic8 - 3.85 %
topic18 - 3.85 %
topic7 - 2.88 %
topic19 - 1.92 %
topic20 - 1.92 %
topic21 - 1.92 %
topic5 - 1.92 %
topic14 - 0.96 %
topic13 - 0.96 %
topic11 - 0.96 %


Bangladesh
topic15 - 28.33 %
topic12 - 13.33 %
topic20 - 10.00 %
topic4 - 8.33 %
topic5 - 6.67 %
topic1 - 6.67 %
topic0 - 3.33 %
topic13 - 3.33 %
topic16 - 3.33 %
topic18 - 3.33 %
topic14 - 3.33 %
topic11 - 3.33 %
topic8 - 3.33 %
topic10 - 1.67 %
topic6 - 1.67 %
United Arab Emirates
topic0 - 19.61 %
topic4 - 11.76 %
topic15 - 11.76 %
topic6 - 7.84 %
topic14 - 7.84 %
topic8 - 5.88 %
topic11 - 5.88 %
topic7 - 5.88 %
topic9 - 3.92 %
topic10 - 3.92 %
topic18 - 3.92 %
topic3 - 3.92 %
topic12 - 1.96 %
topic5 - 1.96 %
topic13 - 1.96 %
topic20 - 1.96 %


Romania
topic0 - 17.46 %
topic4 - 14.29 %
topic15 - 12.70 %
topic12 - 9.52 %
topic20 - 7.94 %
topic16 - 6.35 %
topic8 - 6.35 %
topic7 - 3.17 %
topic6 - 3.17 %
topic17 - 3.17 %
topic18 - 3.17 %
topic14 - 1.59 %
topic3 - 1.59 %
topic10 - 1.59 %
topic2 - 1.59 %
topic5 - 1.59 %
topic13 - 1.59 %
topic11 - 1.59 %
topic21 - 1.59 %
Russian Federation
topic0 - 27.78 %
topic18 - 23.61 %
topic13 - 11.11 %
topic4 - 8.33 %
topic2 - 6.94 %
topic15 - 4.17 %
topic5 - 2.78 %
topic20 - 2.78 %
topic6 - 2.78 %
topic8 - 1.39 %
topic14 - 1.39 %
topic16 - 1.39 %
topic19 - 1.39 %
topic11 - 1.39 %
topic9 - 1.39 %
topic3 - 1.39 %