applied sciences

MDPI

Review

Exploring and Visualizing Research Progress and Emerging Trends of Event Prediction: A Survey

Shishuo Xu [1,2], Jinbo Liu [1,2], Songnian Li [3], Su Yang [4,*] and Fangning Li [2]

1 School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, 1 Zhanlanguan Road, Beijing 102616, China; xushishuo@bucea.edu.cn (S.X.); 2108570021078@stu.bucea.edu.cn (J.L.)
2 Key Laboratory of Urban Spatial Informatics, Ministry of Natural Resources of the People's Republic of China, 15 Yongyuan Road, Beijing 102616, China; lfn96@163.com
3 Department of Civil Engineering, Toronto Metropolitan University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada; snli@torontomu.ca
4 College of Geosciences and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China
* Correspondence: yangsu@student.cumtb.edu.cn

Abstract: Over the last decade, event prediction has drawn attention from both academic and industry communities, resulting in a substantial volume of scientific papers published in a wide range of journals by scholars from different countries and disciplines. However, thus far, a comprehensive and systematic survey of recent literature has been lacking to quantitatively capture the research progress as well as emerging trends in the event prediction field. Aiming at addressing this gap, we employed CiteSpace software to analyze and visualize data retrieved from the Web of Science (WoS) database, including authors, documents, research institutions, and keywords, based on which the author co-citation network, document co-citation network, collaborative institution network, and keyword co-occurrence network were constructed. Through analyzing the aforementioned networks, we identified areas of active research, influential literature, collaborations at the national level, interdisciplinary patterns, and emerging trends by identifying the central nodes and the nodes with strong citation bursts. It reveals that sensor data has been widely used for predicting weather events and meteorological events (e.g., monitoring sea surface temperature and weather sensor data for predicting El Nino). The real-time and multivariable monitoring features of sensor data enable it to be a reliable source for predicting multiple types of events. Our work offers not only a comprehensive survey of the existing studies but also insights into the development trends within the event prediction field. These findings will assist researchers in conducting further research in this area and draw a large readership among academia and industrial communities who are engaged in event prediction research.

Keywords: event prediction; complex network analysis; visual analysis; CiteSpace

Citation: Xu, S.; Liu, J.; Li, S.; Yang, S.; Li, F. Exploring and Visualizing Research Progress and Emerging Trends of Event Prediction: A Survey. *Appl. Sci.* **2023**, *13*, 13346. https://doi.org/10.3390/app132413346

Academic Editor: Andrea Prati

Received: 23 November 2023
Revised: 15 December 2023
Accepted: 16 December 2023
Published: 18 December 2023

1. Introduction

Events occur at a specific location and time with a certain topic, which covers a wide variety of real-world phenomena, ranging from natural disasters (e.g., earthquakes, typhoons, and floods), to public health events (e.g., infectious disease epidemics and adverse drug events), to social security incidents (e.g., crime, terrorist attacks, and economic security incidents). No matter what type of event it is, either event location or event content usually varies over time due to the event evolution mechanism. Through mining the spatiotemporal and semantic evolution patterns from historical events, the occurring time, location, and content of the upcoming events can be predicted. Predicting events with accurate information aids the management department as well as individuals in making decisions in a timely manner.

With the continuous development of data science and artificial intelligence technology, event prediction has drawn attention from scholars of multiple fields, e.g., social science, computer science, statistics, etc. In other words, event detection emerges as an interdisciplinary hotpot with significant characteristics of collaboration among different domains. As such, there appears an increasing number of studies that are related to event prediction in recent years, which makes it difficult to capture the hot topics as well the future trends in a short time.

There are some review papers that survey the event-prediction-related studies [1–7], among which Zhao [4] made the most comprehensive one in the big data era, including the existing techniques, application domains, evaluation procedures, and future directions with regard to event prediction. However, there is a lack of a comprehensive and systematic survey paper that survey and analyze recent literature on event prediction to obtain its evolution patterns over time. With the objective of addressing the aforementioned gap, a few methods can be considered, including bibliometric analysis and scientometric analysis. The bibliometric analysis is focused on organizing and analyzing literature within a specific field to discover the overview and details of this field, which is not conducive to capturing the evolution process, future trends as well as cooperation patterns among interdisciplinary research [8]. The scientometric analysis leverages quantitative methods and indicators to assess the influence and development trends of an author, a document, or a journal in the interdisciplinary research network, helping researchers grasp the hot topics and academic dynamics of the surveyed field [9]. As such, we selected scientometric analysis in this work to survey the event prediction-related literature. Web of Science (WoS) is a comprehensive and widely used online research database that provides access to a vast collection of academic literature, including scholarly articles, conference proceedings, books, and patents, across various disciplines. Web of Science is widely used by researchers, academics, and professionals for literature reviews, citation analysis, and staying up-to-date with the latest research in their respective fields. It is considered a valuable resource for conducting comprehensive and rigorous research across various disciplines. Based on the publications obtained from WoS, the CiteSpace tool of version 6.2.R4, which has been widely used for the scientometric analysis of academic literature, was used to conduct co-occurrence and co-citation network analysis among scholars, publications, institutions, and keywords in order to discover and visualize the central scholars, hot topics, top institutions as well as the future trends in the event prediction field. Those findings are beneficial for scholars to obtain an overview of this field at the frontiers.

The rest of the paper is organized as follows. Section 2 summarizes the background knowledge of event prediction. Section 3 outlines the methodology employed for conducting scientometric analysis in this field. The results are then analyzed, visualized, and discussed in Section 4. Finally, Section 5 draws conclusions based on the findings.

2. Background Knowledge

As the characteristics as well as the evolution patterns of events vary in space, time, and semantics, the predictability of events and how they can be predicted draw concerns from academia and industry. Muthiah, et al. [10] had been running a system for civil unrest event prediction for four years, based on which they systematically summarized the uncertainty of event prediction from two perspectives, i.e., timing and cause. As shown in Table 1, by selecting the known or unknown time and cause as indicators, the events can be classified into four types, including planned events, recurring events, spontaneous events, and black swan events. Since the cause and timing of black swan events are both unknown, there is little research that places focus on predicting such types of events. In the following, we will elaborate and discuss the background knowledge concerning the four types of events.

Table 1. Four types of events with known or unknown timing and cause.

Event Type	Cause	Timing
Planned events	√	√
Recurring events	×	√
Spontaneous events	√	×
Black swan events	×	×

Note: "√" means known and "×" means unknown.

2.1. Planned Event Prediction

The planned events occur with known cause and timing. The time, location, and topics concerning those events are usually announced in advance through social media or other public platforms. Protests organized by political parties, labor, and student unions are the most common type among the panned events. Those who organize the protests would like to post the gathering date, time, and place as well as the specific demands on their websites or social media accounts in order to attract as many people as possible to obtain support. They assume that larger protests tend to be more disruptive and effective at conveying support compared to smaller ones. To mobilize a large number of participants, it is advisable to plan and publicize the protest's details in advance [11–13]. In this case, the planned events can be predicted by detecting the event occurrence and tracking critical indicators from news and social media rather than mining patterns from the previous data. For instance, Basnet, et al. [14] developed a clustering technique based on spatiotemporal k-dimensional structure trees. This method was used to investigate the spatiotemporal patterns of conflict events that occurred in India during 2014 based on the Global Database of Events, Language, and Tone (GDELT) data. Twitter serves as a widely used social media platform that contains abundant information regarding planned events. Iyda and Geetha [15] introduced an Improved Deep Belief Neural Network (iDBNN) to predict protests using Twitter data, where the efficiency of the proposed method was validated with the case study of the 2019 Hong Kong protests. In addition to Twitter, Google Trends (GT) offers a valuable gateway to access extensive big data on various global topics. Timoneda and Wibbels [16] proposed a novel "variance-in-time" approach that utilized GT to predict the protests in the United States, contributing fresh insights into the specific domain of political protests.

2.2. Recurring Event Prediction

Recurring events refer to those events that occur or appear on a regular basis, but the causes may be different. For example, multiple types of protests and violent events in countries with a large number of Muslim people usually take place after the communal prayer on Fridays since there are crowds gathered [17]. Such types of events can be predicted using frequent pattern mining-based methods, time series forecasting-based methods, and temporal point process-based methods.

The frequent pattern mining-based methods predict certain events by exploring the frequent occurrence patterns from a series of historical events. As the type of events to be predicted is determined, the key to this method is to search the associated event sets that frequently co-occur in the previous event streams. An integrated framework was designed to discover the frequent episodes, which were composed by the continuous event sequences, to predict a specific type of events [18,19] or multiple types of events [20]. Zhou, et al. [21] further sorted the predicted events by confidence in a descending order to retrieve the top-k events so as to improve the reliability of prediction results.

The time series forecasting-based methods partition the event streams into a series of events based on constant time intervals, e.g., one hour, one day, or one week. Yonamine [22] took the instability of events in time sequences into consideration and adopted the Auto-Regressive Fractionally Integrated Moving Average (ARFIMA) model to predict the levels of violence in the Afghanistan area. Given a sequence of terrorist incidents data with

different states, Petroff, et al. [23] predicted the future event state by searching for the state that was most likely to generate the sequence data based on the Hidden Markov Model (HMM). A similar approach was also applied for communication system failure detection, showing effective and efficient performance [24]. Air pollution stands as one of the most detrimental environmental issues globally, necessitating the need for effective and accurate prediction of ozone concentrations. Carbo-Bustinza, et al. [25] employed the seasonal trend decomposition method to decompose the time series into three distinct sub-series, i.e., long-term trend, seasonal trend, and random series, for ozone concentration prediction. The methodology of time series prediction is also commonly applied in the realm of studying infectious diseases. Ballı [26] proposed a time series prediction model using machine learning techniques to capture disease curves and forecast epidemic trends.

Different from time series forecasting-based methods that predict events with equal time intervals, the temporal point process-based methods model the events with temporal heterogeneity in order to generate more accurate time stamps. The classical models such as the Poisson point process model [27,28], Hawkes point process model [29], and Weibull point process model [30] for event prediction purposes. As deep learning shows great potential for mining hidden patterns, some scholars have proposed the deep point process model by leveraging neural network technologies for event prediction but with less interpretability [31].

2.3. Spontaneous Event Prediction

The spontaneous events are with known or traceable causes but unknown occurrence times, which make up most of the studies in the event prediction field. With regard to predicting such types of events, the causes can be extracted from crowd-sourcing data using machine learning-based classification methods, unsupervised learning methods, knowledge graph-based methods, and multi-technology fusion methods.

The typical classification methods such as Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM) have been widely used for predicting whether the event occurs (i.e., binary classification problem) or what type of the event (multi-classification problem). Korkmaz, et al. [32] integrated heterogeneous data (e.g., Twitter data, blogs, and currency exchange rates) to predict whether civil unrest would happen based on the LR model. Zhao, et al. [33] proposed a parameter optimization algorithm on the basis of a lasso regression model to predict the event occurrence at a given location during the next time period, which achieved good performance on the datasets of civil unrest events in Brazil and Venezuela. Tama and Comuzzi [34] compared a total of 20 classification models to evaluate their performance for predicting the following event in the business process, where the credal decision tree (C-DT) model performed best.

The unsupervised methods without labeling the data mainly include outlier detection methods and clustering-based methods. Xu, et al. [35] proposed a spatial–temporal-semantic approach to detect local events using geosocial media data. They first extracted spatiotemporal outliers by measuring the geographical regularities of posting tweets, following which the event content was summarized by adopting the topic modeling method. A case study of the 2014 Toronto International Film Festival (TIFF) was conducted and the results illustrated that 87% of the events can be correctly detected. Kattan, et al. [36] proposed a genetic programming-based framework that integrated the k-means clustering analysis method to predict the location of the particular events defined by users in a time series. This approach helped e-marketing managers decide where and when to advertise their products by monitoring and predicting the users' searching trends using the Google Trends data that records the keywords and searching time in different countries. Chen and Neill [37] identified the abnormal clusters in the heterogeneous social media graph to predict the event time, location, type, and participants based on the Non-Parametric Heterogeneous Graph Scan (NPHGS) method. Xu, et al. [38] applied a spatiotemporal clustering-based method to identify traffic events from Twitter data, where the content of detected events was automatically generated using a list of representative terms.

As the knowledge graph emerged in the past ten years, the event knowledge graph has been leveraged for event prediction. The events (or event attributes) compose the nodes and the relationships between events (or between events and their attributes) compose the edges. The causal relationship between events has been mostly investigated through building event causality graphs, based on which the future events are inferred through either knowledge reasoning or semantic web technologies [39–42]. Despite that, other logic relationships such as sequential relationships, conditional relationships, and subordination relationships were also mined for event prediction with the emergence of the event logic graph [43], where a number of neural networks have been recently adopted to learn the embedded event features with the aim of improving prediction accuracy. Chimmula and Zhang [44] combined the Long Short-Term Memory (LSTM) and the Recurrent Neural Network (RNN) model to predict COVID-19 events. Kapoor, et al. [45] further took spatial information into consideration and constructed spatiotemporal graph neural networks to forecast the COVID-19 trends in space and time. With the aim of improving the deep neural networks, Deng, et al. [46] proposed a dynamic graph convolutional network to provide the context of multi-event prediction results.

In addition, some scholars have fused two or more of the above-mentioned methods to predict event time, location, and content. Some researchers developed an EMBERS system integrating five types of prediction models to predict the event time, location, causes, and scale, and the prediction results were sorted by confidence values [13]. A similar system named carbon was proposed by Kang, et al. [47] to predict civil unrest events using news and social media data. Wang, et al. [48] integrated multiple data mining methods for forecasting the extreme flood events occurring in the next 5 to 15 days. Artificial intelligence, human–machine combination, and hybrid intelligence technologies have also been applied for predicting geopolitical events [49,50] and epidemic diseases [51] in order to improve the model performance as well as the human–machine interaction experience.

2.4. Black Swan Events

Black swan events are rare, unexpected, unpredictable, and highly influential [52,53]. In terms of risk management and decision-making under uncertainty, much attention has been paid to so-called black swans [54,55], such as the financial crisis in 2008 [56], the 911 terrorist attacks in 2001 [57] and the COVID-19 event [58]. Unpredictable extreme weather events often have particularly severe consequences as well [59]. Such types of events have resulted in long-term influence on the globe. Due to the extreme nature of black swan events, they often fall outside the range of normal or conventional events and lack historical records for reference. In this case, the black swan events are unpredictable since most of the event prediction models are constructed and trained with historical data [60]. As such, there exists little literature that focuses on predicting black swan events in recent studies.

In order to provide an overview of the aforementioned types of events that have been mainly selected for predictive study, we further summarized the subclasses (i.e., scenario cases) of planned events, recurring events, and spontaneous events by reviewing a list of relevant literature [61–65] in Table 2, where a set of keywords regarding each type is also presented.

Table 2. A summary of event types and relevant keywords.

Event Type	Classification Category	Keywords
Planned events	Political activities, performance, sports events, celebrations	Political polls, voting behavior, political forecasting, political campaigns, political sentiment and trends, artistic performance, concert, theater, sports prediction, holiday trend, celebration event, holiday culture

Table 2. *Cont.*

Event Type	Classification Category	Keywords
Recurring events	Seasonal weather events, regular update and maintenance of the computer, traffic congestion, regular meeting, religious ceremony, financial events regularly	Seasonal precipitation, seasonal climate changes, meteorological seasons, seasonal weather, seasonal patterns, recurring events, system maintenance, patch management, preventive maintenance, IT infrastructure, commuter patterns, transportation planning, rush hour, commuter behavior, public transportation, meeting scheduling, meeting frequency, agenda setting, worship practices, rituals and traditions, ceremonial practices, faith-based celebrations, religious festivals, financial events, economic cycles, market fluctuations
Spontaneous events	Geological hazards, sudden network attacks and data leaks in computers, traffic accident, unexpected weather events, emergency health crisis, sudden environmental disaster (oil spill, etc.)	Earthquake prediction, volcanic activity, landslide susceptibility, natural disaster, geological monitoring, geological hazards, cyberattacks, intrusion detection, cyberattacks, network security, hacker attacks, traffic collisions, vehicle safety, road infrastructure, driver behavior, emergency response, traffic accident, extreme weather, tornado outbreaks, severe storms, climate anomalies, hurricane, typhoon, public health emergency, epidemic, pandemic, health crisis management, environmental crisis, pollution incident, ecological impact, environmental monitoring

3. Methodology

Figure 1 illustrates the overall procedure for conducting the scientometric analysis of the publications in the field of "event prediction", including data collection, author co-citation network analysis, document co-citation network analysis, collaborative institution network analysis, and keyword co-occurrence network analysis. In this way, the collaboration patterns among authors and institutions as well as the dynamics of the domain can be identified.

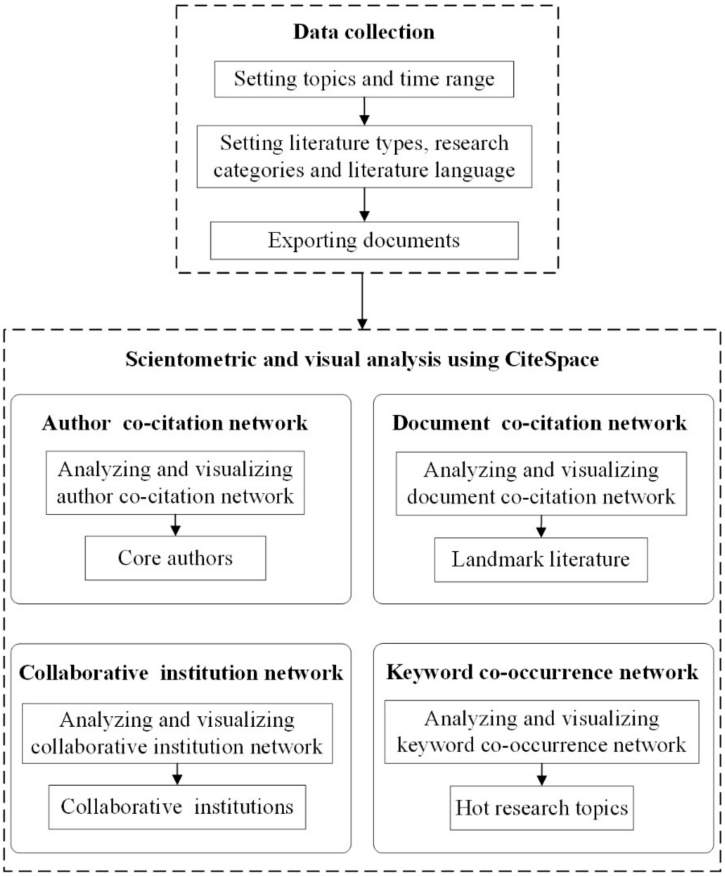

Figure 1. The overall framework of analyzing relevant event prediction research in a scientometric way.

3.1. Data Collection

We collected literature from the Web of Science (WoS) website on February 11, 2023, to prepare data for scientometric analysis. To begin with, we used (TP = (event * prediction) OR (event * forecasting)) as topics in the WoS to perform a broad search, where "OR" means at least one topic appearing in the document, and "*" means any characters. In order to investigate how event prediction research has evolved over the last decade, which is usually selected as the time interval for conducting scientometric surveys in multiple disciplines, the time range was set from 1 January 2012 to 31 December 2022. As a result, a total of 9191 documents were returned. Since the documents with the selected topic anywhere in the full paper can be returned, it is likely that a large number of documents are not closely related to event prediction in the original search results. We added further filters by taking three types of constraints into consideration. First, the literature types were set as articles, review papers, editorial materials, and conference papers. Second, the research category was set as Meteorology Atmospheric Sciences, Computer Science Information Systems, Environmental Sciences, Geosciences Multidisciplinary, Computer Science Artificial Intelligence, Computer Science Interdisciplinary Applications, Imaging Science Photographic Technology, and Transportation Science Technology, which covers almost all event types summarized in Section 2 as well as retains effective sources for analyzing interdisciplinary research patterns. Third, the language was set as English. As a result, a total of 4473 documents were finally obtained and exported.

3.2. Scientometric and Visual Analysis Using CiteSpace

A number of tools can be used for scientometric and visual analysis of the literature in a research field, including Bibexcel, Gephi, VOSviewer, and CiteSpace. A summary of their advantages and disadvantages is presented in Table 3.

Table 3. A summary of scientometric analysis tools.

Tool	Advantages	Disadvantages
Bibexcel	• Small and practical • Strong compatibility with other software	• Requiring specific input data format • Poor visualization ability • Lack of dynamic analysis
Gephi	• Large user community • Supporting multiple data format • Strong visualization ability	• Cumbersome operation of data cleaning and data conversion • Limited functions of analyzing scientific literature • Lack of quantitative indicators for explaining results
VOSviewer	• User-friendly operation interface • Supporting geographic visualization • Supporting multiple data format	• Poor visualization effect of cluster analysis results • The node information of the visual network cannot be viewed
Citespace	• Powerful visualization function based on various criteria • Capable of analyzing large-scale networks • Rich functions for scientific literature analysis • Supporting dynamic analysis	• Requiring high computer performance • The old version of the software cannot be used after the new version of the software appears

With the aim of visualizing and analyzing multiple networks to capture hot topics and development trends of event prediction research, we finally chose to use CiteSpace software [9] by importing those literature documents collected from the WoS in a scientometric way. CiteSpace is an academic literature visualization analysis software developed on the basis of data visualization and metrology, which specializes in analyzing potential knowledge in scientific research. Since the software analyzes the structure, rule, and distribution of scientific knowledge presented by means of visualization, the analysis results are called "scientific knowledge map" [66].

Specifically, the deduplication function of CiteSpace was first used to remove duplicate documents and retain a single source, after which 4438 valid documents were left. As shown in Figure 2, we counted the number of documents published year by year. It reveals

that an overall upward trend occurs, especially since 2020, the number of published papers on event prediction has increased significantly. Among all the documents, 94.9% are articles, 2.7% are review papers, 2.2% are conference papers, and 0.2% are editorial materials. The distribution of the literature types is shown in Figure 3.

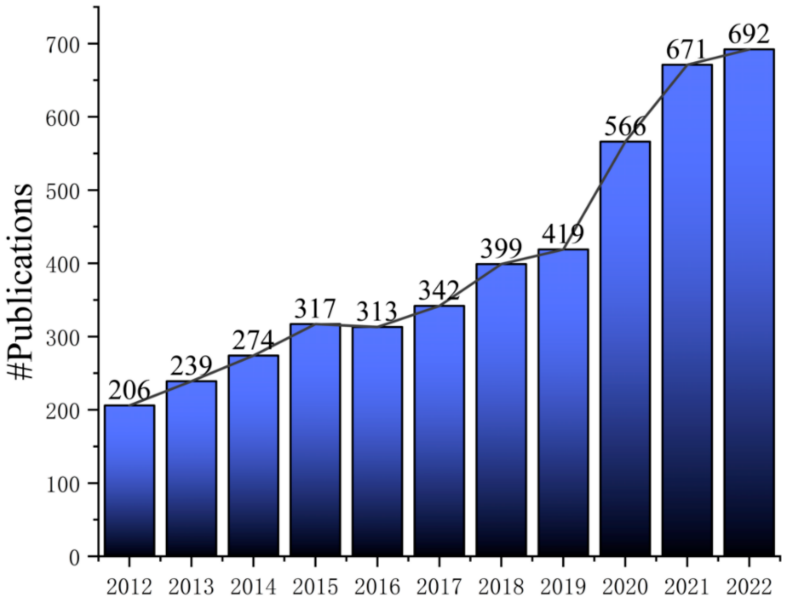

Figure 2. The number of publications on "event prediction" for the years 2012–2022.

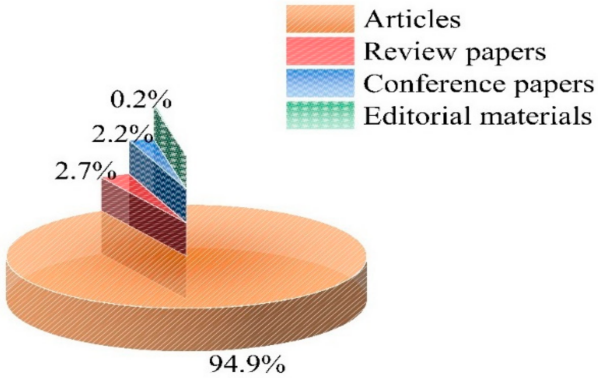

Figure 3. The distribution of the literature types for the years 2012–2022.

Based on the preprocessed documents, we used CiteSpace software to conduct four types of analyses, including author co-citation network analysis, document co-citation analysis, collaborative institution network analysis, and keyword co-occurrence network analysis. Specifically, the co-citation networks of authors and documents aimed at identifying the influential academic researchers and papers, respectively. The collaborative institution network was composed to identify cooperation patterns among different institutions. Finally, we analyzed the keyword co-occurrence patterns and the timeline of keywords to obtain the development of research hotspots. The key parameters as well as parameter settings involved in analyzing and visualizing the four types of research networks are introduced as follows.

With regard to the networks generated by CiteSpace, they are composed of nodes and lines. The nodes refer to the objects to be analyzed (e.g., authors and documents), and the lines represent the interaction between nodes. The thicker the line is, the stronger the relationship between the nodes is. The size of nodes indicates the importance of the objects in the network, which can be measured by a variety of indicators, such as citation frequency and the number of publications. The concentric ring of different colors enclosing

the nodes represents the time span information of the nodes. Each color ring represents a specific year. The networks can be analyzed and visualized based on degree centrality, betweenness centrality, closeness centrality, citation frequency, or citation burstiness.

Centrality measures the importance of a node within a network in terms of its connections, interactions, and influence on other nodes. In this paper, degree centrality, betweenness centrality, and closeness centrality were used for centrality measurement. The degree of centrality quantifies the number of direct connections a node has in the network. Nodes with higher degrees are regarded as more central due to their larger number of immediate connections. The betweenness centrality is used to quantify the degree of mediation of nodes in the network, that is, the importance of nodes in the shortest path between different nodes. By calculating the betweenness centrality of nodes, we can identify the central nodes in the network. In CiteSpace, the betweenness centrality score ranges between 0 and 1. Nodes with a betweenness centrality that is not less than 0.1 are highlighted with purple concentric rings. The thickness of the purple rings is proportional to the betweenness centrality score. The closeness centrality is calculated by taking the reciprocal of the sum of the shortest path lengths from a node to all other nodes in the network. The higher the closeness centrality value, the more central the node is in terms of its proximity to others. Citation frequency reflects the number of times a node is cited within a specific time period, providing insights into the influence and popularity of that node in the field of event prediction. Nodes that experience a substantial increase in citations within a specific time frame are identified as "burst" nodes. The occurrence of citation bursts allows researchers to discern trends in research development.

Furthermore, we performed cluster analysis for the author co-citation network, document co-citation network, and keyword co-occurrence network, where two metrics including modularity and silhouette value were used to explain the scholarly framework of the event prediction domain. Modularity is a metric that quantifies the extent to which a network can be partitioned into distinct components or modules. The modularity measures the degree of connection between nodes in the same module divided by the degree of connection between nodes in different modules, which is usually measured using the Q value. If a network has a high modularity (i.e., the Q value is larger than 0.3), it means that there is an obvious modular structure in the network. The silhouette value, also known as the S value, which ranges from -1 to 1 [9], is another metric used to assess the quality of a clustering configuration. A higher silhouette score indicates a higher level of homogeneity within the cluster and a higher level of heterogeneity among clusters. The S value of 0.7 is often taken as the threshold to determine whether the clustering results are reasonable. With the purpose of interpreting the clustering results, we adopted the log-likelihood ratio (LLR) algorithm for label analysis methods, which generates a set of representative terms to explain the cluster content. LLR measures the degree of association by comparing the observed and expected frequencies of words (or terms) in a document. The specific calculation formula is as follows:

$$\mathrm{LLR} = \log \frac{p(C_j \backslash V_{ij})}{p(\overline{C}_j \backslash V_{ij})},$$

where $p(C_j \backslash V_{ij})$ and $p(\overline{C}_j \backslash V_{ij})$ refer to the density function of the feature vector V_{ij} in the cluster C_j and $\overline{C}_j$, respectively. The larger the LLR, the more representative the word is to the cluster. The words with high LLR values serve as the chosen labels for the cluster.

4. Results and Discussion

4.1. An Overall View of the Types of the Predicted Events

According to the information summarized in Table 2, we identified and classified the events predicted in the obtained 4473 documents through keyword matching. As shown in Figure 4, around half the events belong to "spontaneous events", which indicates that predicting those events is of great significance and draws a large number of scholars

to get involved in such research. The documents related to "planned events" hold the smallest number. It is likely that such events with planned time, location, and content are less appealing to prediction research compared to those events with uncertainty, such as spontaneous events and recurring events. The distribution of the number of documents related to each event type is almost consistent with the elaboration in Section 2.

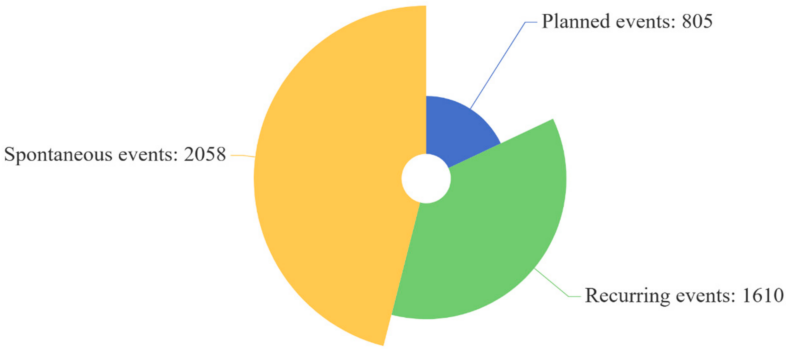

Figure 4. The distribution of the number of documents related to each event type.

4.2. Author Co-Citation Network Analysis

This section assesses the influential authors in the field of event prediction by conducting the author co-citation network analysis in CiteSpace. The analysis encompasses the period from 2012 to 2022, with a particular emphasis on the top N authors within each one-year segment. The N usually falls between 10 and 50. A large value of N poses great pressure for CiteSpace for network analysis and visualization that returns results with low efficiency, while a small value leads to the sparse relationship between nodes and certain research patterns are likely to be missed in this case. In order to keep the balance between computation efficiency and result integrity, we set the N as 20 in this study. This number is typically chosen for network construction in the majority of scientometric analysis studies [67–69]. The anonymous authors were excluded from the analysis. As shown in Figure 5, the author's co-citation network includes 116 nodes and 645 links. The nodes in the network represent individual authors. When two authors are cited in the same document, a link is established between them, indicating a co-citation relationship. Each color in the visualization corresponds to a specific time slice, typically one year. The concentric rings display the changing patterns of author co-citations over time, with different colors representing different periods.

In Figure 5, the larger the node size is, the more frequently cited the author is. The nodes (i.e., authors) owning high citation frequency can be regarded as the core authors in the event prediction field. It reveals that Fausto Guzzetti has the highest citation frequency of 263 during each time slice, whose node holds the largest radius in the whole network, followed by Leo Breiman and Dick Dee with 185 citations and 170 citations, respectively. It reveals that their publications related to event prediction are much more popular and acknowledged by scholars in the relevant fields.

The purple rings enclosing the concentric circles represent the betweenness centrality of the author. The thicker the purple ring, the higher the betweenness centrality. Table 4 lists the top five authors in the event prediction field sorted by the betweenness centrality scores. It shows that there is no significant difference among those authors, indicating that they play critical roles in connecting all authors to compose the research network. Specifically, William C. Skamarock owns the highest betweenness centrality and closeness centrality, he is from the National Center for Atmospheric Research (NCAR), United States, and his research interests cover a few areas, including Meteorology and Atmospheric Sciences, Geology, Physics, Computer Science, and Environmental Sciences and Ecology. It reveals that he has been a promoter of cooperative and interdisciplinary research as well as an important author with the sum of the shortest path from an author to all other authors in the event prediction field since the year 2013. Fausto Guzzetti holds the lower betweenness

centrality but has the highest degree of centrality, indicating that he has the most direct connections with 45 authors to compose the partial network but is less important for connecting all authors to compose the entire network.

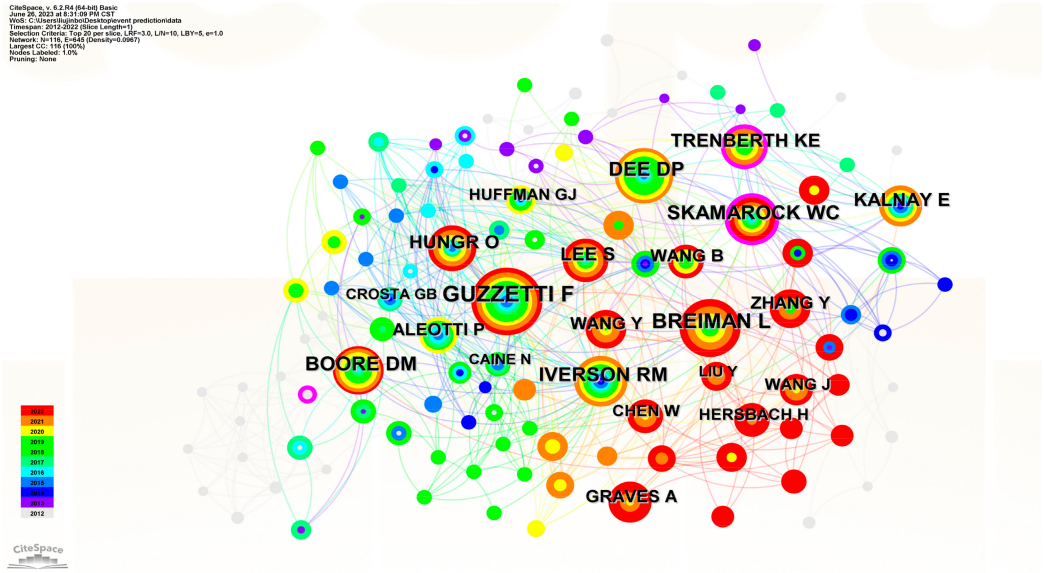

Figure 5. The author's co-citation network for the years 2012–2022.

Table 4. The top five authors sorted by the betweenness centrality.

Author	Betweenness Centrality	Degree Centrality	Closeness Centrality	Year
William C. Skamarock	0.12	33	0.83	2013
Kevin Edward Trenberth	0.11	27	0.76	2012
Eugenia Kalnay	0.10	26	0.65	2012
Richard M. Iverson	0.09	12	0.47	2012
Fausto Guzzetti	0.09	45	0.71	2012

We summarized the top five authors by ranking their citation burst scores in Table 5. The highest score was obtained by Alex Graves with a citation burst of 22.08. Alex Graves has 57 publications and 92,987 citations on the WoS. His event prediction-related studies have been increasingly gaining attention from scholars since 2020 and reached an h-index of 34 on the WoS, which is a metric usually used for measuring a research's academic achievement. The other three authors Wei Chen, Samuele Segoni, and Dieu Tien Bui exhibit similar citation burst patterns over time, indicating that event prediction has emerged as a prominent research direction. This observation suggests that an increasing number of scholars have directed their attention towards this field in recent years.

Table 5. The top five authors sorted by the citation burst.

Author	Citation Burst	Year (Begin to End)
Alex Graves	22.08	2020–2022
Wei Chen	15.66	2020–2022
Samuele Segoni	15.15	2020–2022
Dede Sinan Akkar	15.07	2017–2019
Dieu Tien Bui	14.72	2020–2022

We further conducted a cluster analysis to detect author clusters exhibiting comparable co-citation patterns in the event prediction field. Those authors with similar academic influence and cooperative relationships are grouped in one cluster. As presented in Figure 6, there exist five significant clusters in the network, which are rendered in different colors.

The modularity indicated by the Q value equals 0.5069 which is over 0.3, reflecting that the modularization of the author co-citation network is significant. The authors in the same cluster are homogeneous, which can be distinguished from other clusters since the average S value equals 0.786 which is larger than 0.7, proving the clustering outcomes are both reasonable and desirable.

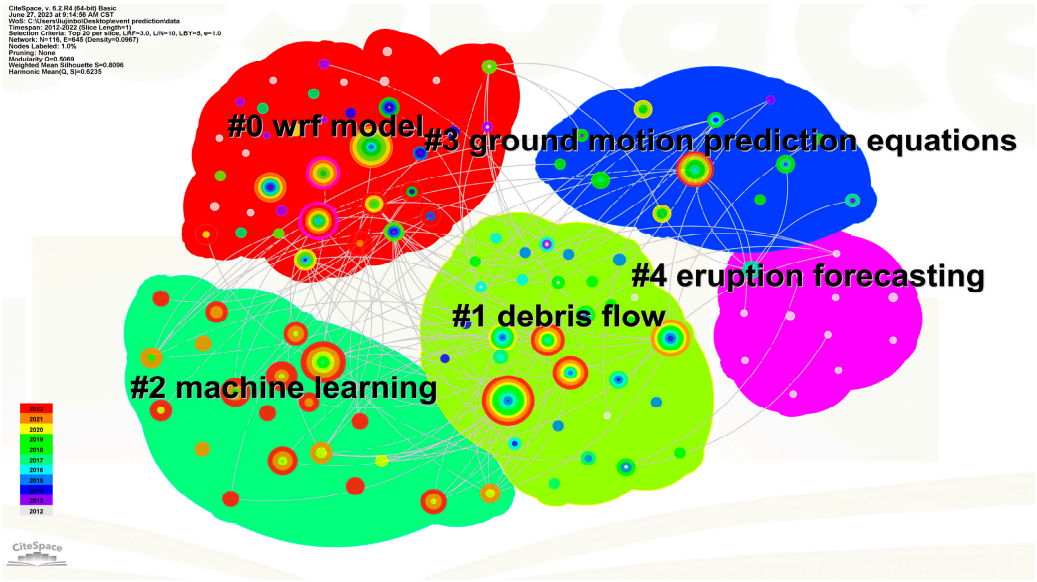

Figure 6. The clusters in the author co-citation network for the years 2012 to 2022.

Table 6 presents the cluster details, arranged in descending order based on the cluster size, which refers to the number of authors within each cluster. The table also includes the S value, mean year, and labels generated by the LLR method for each cluster, aiding in the interpretation of the clustering outcomes. Notably, all S values surpass 0.7, indicating the meaningful grouping of authors with similar research interests into distinct clusters. For example, cluster #0 appearing around the year 2013 includes 37 authors, who were mainly engaged in predicting and validating weather-related events, e.g., extreme sea level forecasting and performance assessment by regions. Other clusters concerning rainfall events and debris flow (cluster #1), ground motion events (cluster #3), and eruption and earthquake (cluster #4) occur in similar years to cluster #0. In the most recent years, deep learning technologies such as convolutional neural network has rapidly emerged in the event prediction field, which aligns with the research trends of leveraging artificial intelligence (AI) for academic research as well as industrial applications.

Table 6. The largest five clusters in the author co-citation network.

ID	Size	Silhouette	Mean (Year)	Label (LLR)
0	37	0.758	2013	Weather research
1	31	0.83	2014	Rainfall threshold; Debris flow
2	20	0.777	2019	Deep learning; Convolutional neural network
3	13	0.794	2015	Ground motion model; Ground motion prediction equation
4	12	0.946	2012	Eruption forecasting; Earthquake forerunner

4.3. Document Co-Citation Network Analysis

In this section, we conducted a document co-citation analysis to determine the key literature from 2012 to 2022. Similar to selecting the top N nodes for network analysis

in Section 4.2, we selected the top 20 documents during each time slice for document co-citation network analysis. The merged network includes 283 nodes and 469 links as shown in Figure 7. The nodes represent independent documents, each of which is indicated by the author and publication time. The links refer to the co-citation relationship between documents. Similar to the author's co-citation network, the concentric rings of different colors in the document co-citation network reflect the co-citation patterns of the documents over time. The largest node in the network represents a paper published by Hans Hersbach (2020) with the highest citation frequency of 30 times. The orange and red thick rings enclosing this node indicate that this paper has been frequently cited and received extensive attention from scholars in 2021 and 2022, aligning with the color bar shown at the left bottom of Figure 6.

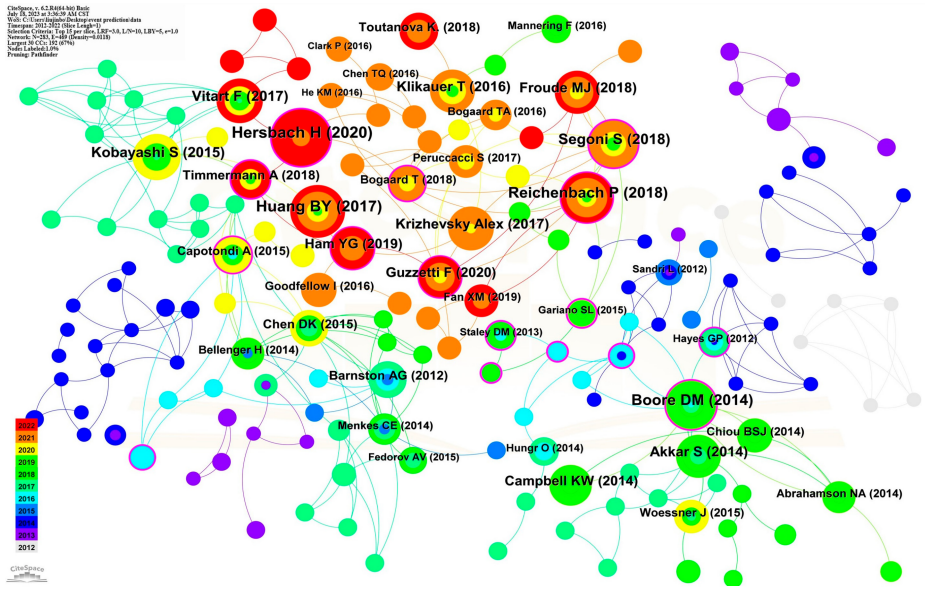

Figure 7. The document co-citation network for the years 2012–2022.

The outermost ring surrounding the node represents the betweenness centrality of the document, indicated by the color purple. The thicker the purple ring, the higher the betweenness centrality of the document. The information regarding the top five documents with higher betweenness centrality in the document co-citation network is shown in Table 7, including title, betweenness centrality, degree centrality, closeness centrality, authors, year of publication, and published journal. The betweenness centrality scores of those documents are all around 0.40 and published in highly impacted journals, covering various topics such as forecasting shallow landslides, determining rainfall intensity and duration, predicting traffic flow, reviewing social media-based event detection techniques, and predicting human behaviors. Specifically, the paper entitled "Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy" which was published in Geomorphology in 2015 stands out with the highest betweenness centrality of 0.45 and highest degree centrality, reflecting this paper plays an important intermediary role and serves as a bridge connecting different research groups in the event prediction field. Similarly, there is a slight difference among documents regarding closeness centrality. The journal paper "Deep learning for short-term traffic flow prediction" published in Transportation Research Part C: Emerging Technologies in 2017 has the highest closeness centrality of 0.67, indicating that this paper plays an important role in connecting all documents through the shortest path in the network.

Table 7. The top five documents sorted by betweenness centrality.

Title	Betweenness Centrality	Degree Centrality	Closeness Centrality	Author	Year	Source
Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, Southern Italy	0.45	22	0.64	Gariano, S. L. et al.	2015	Geomorphology
Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in Southern California	0.44	20	0.65	Staley, D. M. et al.	2013	Landslides
Deep learning for short-term traffic flow prediction	0.44	16	0.67	Polson, N. G., and Sokolov, V. O.	2017	Transportation Research Part C: Emerging Technologies
Review on event detection techniques in social multimedia	0.40	13	0.61	Garg, M., and Kumar, M.	2016	Online Information Review
Ontology-based deep learning for human behavior prediction with explanations in health social networks	0.39	16	0.63	Phan, N. et al.	2017	Information Sciences

When a document experiences a substantial increase in the number of citations within a specific timeframe, it is recognized as having a pronounced citation burst. The paper titled "The JRA-55 reanalysis: general specifications and basic characteristics" was proposed by the Japan Meteorological Agency (JMA) who performed the second Japanese global atmospheric reanalysis (JRA-55) from 1958, gaining the strongest citation burst of 8.03 (see Table 8). The potential reason is that this milestone paper was frequently cited for analyzing and forecasting the meteorological events during 2019 and 2020. Another strongly cited document in the most recent years (i.e., from 2019 to 2022) is a review paper surveying the rainfall thresholds for landslide occurrence since threshold selection is usually a big concern among scholars and this paper provides highly persuasive reference. The other documents with strong citation bursts mainly aim at proposing specific equations and models for predicting meteorological events and natural hazards. Their citation bursts last for two or three years during 2015 and 2019, indicating during those periods, those equations and models had gained widespread recognition and acceptance within the academic community.

Table 8. The top five documents sorted by the citation burst.

Title	Citation Burst	Author	Burst Year (Begin to End)	Source
The JRA-55 reanalysis: general specifications and basic characteristics	8.03	Kobayashi, S. et al.	2019–2020	Journal of the Meteorological Society of Japan
A review of the recent literature on rainfall thresholds for landslide occurrence	7.66	Segoni, S. et al.	2019–2022	Landslides
NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes	7.58	Boore, D. M. et al.	2016–2019	Earthquake Spectra
Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East	6.86	Akkar, S. et al.	2017–2019	Bulletin of Earthquake Engineering
Skill of real-time seasonal ENSO model predictions during 2002–2011: Is our capability increasing	6.26	Barnston, A. G. et al.	2015–2017	Bulletin of the American Meteorological Society

To identify distinct groups of documents with high homogeneity, we performed cluster analysis on the document co-citation network. Figure 8 illustrates the network divided into five clusters, each represented by a different color. The modularity, indicated by a

Q value of 0.8537, and the weighted S value of 0.932 both fall within desirable ranges. These cluster analysis results provide reliable insights that align with our expectations, enabling scholars to gain a better understanding of the academic structure surrounding event prediction-related research.

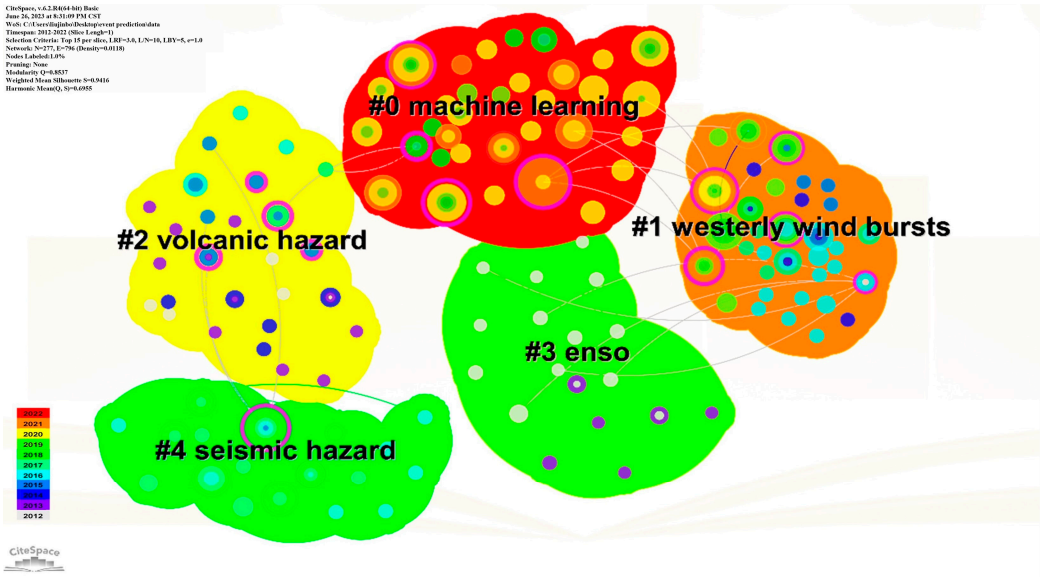

Figure 8. The clusters in the document co-citation network for the years 2012 to 2022.

The details of the five clusters are illustrated in Table 9, including the number of documents, S value, mean year, and cluster labels generated by the LLR method. Cluster #0 and cluster #1 include the same number of documents but take place in different years. The topic of cluster #0 concerns machine learning technologies leveraged for event prediction in recent years. The other clusters all occurred before 2015. Cluster #1 and cluster #3 talk about climate phenomena by exploring their evolution patterns, e.g., seasonal changes. Cluster #2 and cluster #4 focus on monitoring and predicting natural disasters such as seismic hazards and seismic hazards and sensitivity analysis was also conducted to optimize the prediction models.

Table 9. The largest five clusters in the document co-citation network.

ID	Size	Silhouette	Mean (Year)	Label (LLR)
0	33	0.993	2017	Machine learning
1	33	0.878	2014	Westerly wind bursts; El Nino diversity
2	27	0.923	2012	Volcanic hazard; Volcanic eruption
3	18	0.870	2013	Enso; Seasonal prediction
4	18	0.996	2014	Seismic hazard; Sensitivity analysis

4.4. Collaborative Institution Network Analysis

In this section, we examined and visualized the collaboration patterns among research institutions worldwide by constructing a collaborative institution network. The institution information was extracted from the authors' affiliations, resulting in a network consisting of 141 nodes and 868 links. Each node represents an institution, while the links indicate cooperative relationships between institutions. The size of each node corresponds to the frequency of event prediction-related publications from that institution. Figure 9

showcases the node representing Centre National de la Recherche Scientifique in France, which possesses the largest size, indicating a publication frequency of 195. Similarly, the Chinese Academy of Sciences in China also has a node with a comparable size, reflecting a publication frequency of 192. The contributions of these two institutions have significantly advanced the field of event prediction.

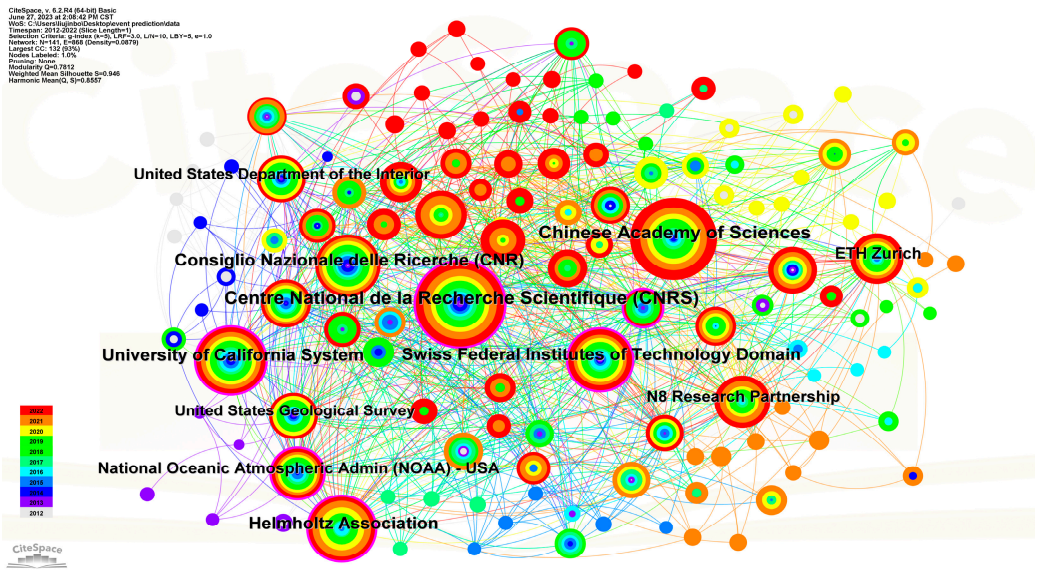

Figure 9. The collaborative institution network for the years 2012–2022.

The concentric rings encircling the nodes display the temporal patterns of documents published by each institution. The color of the links corresponds to the year when the collaboration between institutions first emerged. The thickness of the outermost purple ring signifies the significance of the institution in maintaining interconnected relationships within the collaborative institution network. Institutions with thicker purple rings have higher betweenness centrality scores, indicating their crucial role in bridging connections among all institutions involved in event prediction research. Table 10 illustrates the top five institutions measured by the betweenness centrality. It reveals that the Centre National de la Recherche Scientifique, France, which has published the most event prediction relevant articles since 2012, also has the highest betweenness centrality, degree centrality, and closeness centrality. As such, it can be identified as the most core research institution with significant cooperation advantages and can communicate and cooperate with others more quickly in the event prediction field. The University of California System and the National Oceanic Atmospheric Admin in the United States of America and the Helmholtz Association in Germany have the same centrality score of 0.15, reflecting that they are in the collaborative institution network.

In order to investigate the developmental trends of event prediction-related research in the last decade, we performed a citation burst analysis focusing on research institutions. The top ten institutions with strong citation bursts are listed in Figure 10, including the strength score as well as the beginning year and the end year. The National Taiwan University, Swiss Federal Institute for Forest, and University Toulouse III—Paul Sabatier have the highest citation burst scores in the very beginning years, covering a time span of around two years. During the mid-term, the strongest citation burst appeared in Hohai University, China with a strength of 6.34 and lasted a period of three years from 2014 to 2017. Kyoto University and Duy Tan University have received a lot of attention from other researchers in the most recent years.

Table 10. The top five institutions sorted by the betweenness centrality.

Institution	Betweenness Centrality	Degree Centrality	Closeness Centrality	Country	Year
Centre National de la Recherche Scientifique	0.17	95	0.93	France	2012
University of California System	0.15	75	0.79	United States of America	2012
Helmholtz Association	0.15	72	0.87	Germany	2014
National Oceanic Atmospheric Admin	0.15	53	0.84	United States of America	2012
Swiss Federal Institutes of Technology Domain	0.12	65	0.86	Switzerland	2012

Figure 10. The citation burst history of the institutions in the timespan of 2012–2022.

4.5. Keyword Co-Occurrence Network Analysis

Keywords serve as a valuable means to succinctly encapsulate the key topics covered in a document. The keyword co-occurrence network analysis was performed to observe the connections and development of research topics in the event prediction field. According to the elaboration of selecting N terms for composing a network in Section 4.2, we selected the top 20 keywords per year in the last decade for analysis. As revealed in Figure 11, the network consists of 328 nodes and 752 links. The nodes represent the distinct keywords. If the two keywords appear in one document at the same time, a link is built between the two keywords to illustrate their co-occurrence relationship. The color of the link indicates the year in which the co-occurrence relationship initially emerged. The node size indicates how often the keyword has been used in the surveyed documents. It can be seen in Figure 11 that the node of "extreme event" holds the largest size with a publication frequency of 74, indicating that predicting extreme events has drawn the most interest and appealed to those researchers to get engaged in the multidisciplinary studies. In addition, the keywords "neural work", "deep learning", "machine learning", "numerical weather prediction" and "el nino" also occur frequently. Neural networks, deep learning and machine learning acting as advanced technologies in computer science and data science have been frequently used for event prediction. This aligns with the related work reviewed in Section 2 that deep learning and machine learning methods have been applied for predicting spontaneous events and recurring events in recent years. As indicated in Table 2, the unexpected weather event is a typical type of spontaneous event, which is consistent with the results that the keywords "numerical weather prediction" and "el nino" frequently occur in the keyword co-occurrence network. In a word. the frequent occurrence of the aforementioned keywords reflects the diversity and complexity of cross-domain event prediction research, where researchers attempt to address the problems they face by adopting advanced techniques and multidisciplinary approaches.

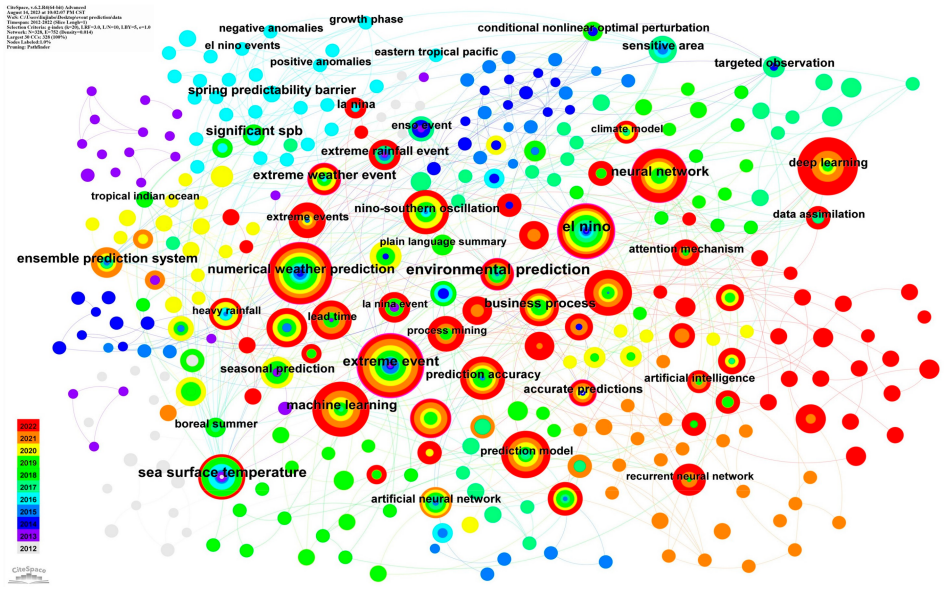

Figure 11. The keyword co-occurrence network for years 2012 to 2022.

In Figure 11, the purple rings outmost the concentric circles reflect the betweenness centrality of the keywords. The thicker the purple rings, the higher the betweenness centrality of the keywords. A high betweenness centrality score signifies that the associated keyword acts as a crucial intermediary in connecting all co-occurring keywords within the keyword co-occurrence network. The top five keywords with the highest betweenness centrality are displayed in Table 11. The keywords "extreme event" and "el_nino" with higher publication frequency also hold the highest betweenness centrality, degree centrality as well as closeness centrality, revealing that predicting extreme events, especially weather-related events (e.g., "el_nino" and "extreme weather event") have been hot research focus and of great significance in the event prediction field during the past decade and play important roles in the partial network as well as the entire network. Predicting extreme weather events requires monitoring weather data such as atmospheric conditions, temperature, humidity, etc., which is usually collected through weather sensors and weather stations. The El Nino prediction usually requires monitoring climate and meteorological indicators such as sea surface temperature from satellites and ocean sensors. Evaluating the performance of prediction models and methods with accuracy is also a concern in the event prediction field.

Table 11. The top five keywords sorted by the betweenness centrality.

Keywords	Betweenness Centrality	Degree Centrality	Closeness Centrality	Year
Extreme event	0.23	87	0.86	2012
El Nino	0.16	64	0.77	2013
Sea surface temperature	0.13	21	0.68	2012
Extreme weather event	0.10	29	0.73	2016
Prediction accuracy	0.08	38	0.69	2012

Table 12 presents the top five keywords with the strongest citation bursts. A stronger citation burst is indicated by a higher strength score. The term "recurrent neural network" exhibits the most robust citation burst, with the highest score of 13.32 during 2017 and 2022. This is because recurrent neural network as a typical type of deep learning method has been widely used for event prediction in multidisciplinary fields, aiming at improving the performance of prediction models with advanced techniques. The "extreme event" with a citation burst of 11.32 has received an increase in attention since 2018.

Table 12. The top five keywords sorted by the citation burst.

Keywords	Citation Burst	Year (Begin to End)
Recurrent neural network	13.32	2017–2022
Prediction model	12.39	2016–2022
Extreme event	11.32	2018–2022
Weather sensor	10.07	2014–2022
Artificial intelligence	7.35	2019–2022

Sensors have usually provided effective data sources for monitoring and predicting abnormal situations (i.e., events) in multiple fields since 2014, especially in dealing with weather event prediction. For instance, weather sensor data can be imported into models such as recurrent neural networks to capture the association between meteorological conditions and extreme weather events. Another keyword cited with a significant increase in the most recent years is "artificial intelligence", of which the citation burst score is 9.75. It is a trend that the large-scale collection and processing of sensor data require powerful computing power and intelligent algorithms.

Furthermore, we constructed a two-dimensional keyword timeline graph in Figure 12, which arranges the keyword co-occurrence network and the corresponding cluster analysis results in chronological order, so as to investigate the distribution of research topics in the event prediction field and how they evolve over time. The horizontal axis on the top signifies time. Each horizontal line denotes a cluster. The clusters are ordered according to the earliest occurrence time of the documents in each cluster. The circle located on the horizontal line represents the keywords in the cluster, and the position of the circle is the time when the keyword first emerged in the cluster. The concentric rings around each keyword illustrate the time range during which the keyword appeared.

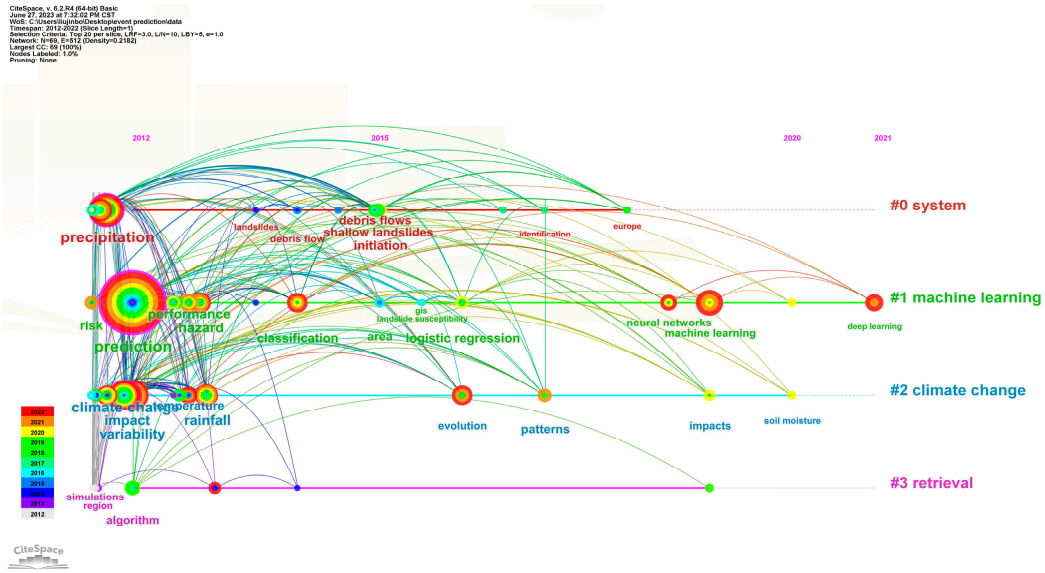

Figure 12. The keyword timeline graph for the years 2012–2022.

The labels of the clusters were generated by the LLR method. Cluster #0 was labeled as "system", where the representative keywords include "precipitation", "debris flows", "landslides" and "shallow", revealing that a number of event prediction systems were built for various purposes. The "machine learning" was assigned cluster #1, where a list of model representations such as "logistic regression" and "neutral networks" was adopted in the event prediction field. The deep learning methods emerged as a hot keyword in 2021. Another research focus in the event prediction field was placed on climate change. Scholars investigated the evolution patterns, variability, and impact of temperature, rainfall, and

soil moisture with the prediction purpose. Cluster #4 labeled as "retrieval" illustrates that the simulation algorithms were used for event prediction between 2012 and 2018.

5. Conclusions

This study utilizes CiteSpace software to conduct a scientometric analysis of research focused on event prediction. To explore research productivity and emerging trends in the surveyed field, we gathered documents published from 2012 to 2022 from the Web of Science database. The number of publications concerning event prediction has kept increasing in the last decade, especially with significant increases since 2019, indicating this topic has attracted more and more attention from the research communities over time. Four networks were then generated and visualized based on the collected documents for scientometric analysis, including the author co-citation network, document co-citation network, collaborative institution network, and keyword co-occurrence network. The four types of network analysis have produced several remarkable findings. The representative authors and documents are from different disciplines (e.g., meteorology science, computer science, and data mining), showing that event prediction research holds obvious interdisciplinary characteristics promoting cooperation and communication among research institutions worldwide. The frequently co-occurring keywords reveal hot research topics (e.g., extreme events and weather events) and widely used methods (e.g., deep learning and machine learning methods) as well as the research trends (e.g., artificial intelligence aiding event prediction) with great potential. In addition, sensor data has been playing an integral role in event prediction since 2014, especially for predicting weather events and meteorological events (e.g., monitoring sea surface temperature and weather sensor data for predicting El Nino). The real-time and multivariable monitoring features of sensor data enable it to be a reliable source for predicting multiple types of events.

In spite of the notable results achieved in this work, there is still room for improvement in the near future. The scope of this study was limited to articles written in English and collected exclusively from the Web of Science database. As a result, it is possible that certain pertinent research might have been inadvertently omitted from this survey. Aiming at overcoming this limitation, the documents written in other languages (such as Chinese) and literature data obtained from other databases (such as Scopus) can be incorporated to enable a more comprehensive comparison and analysis in the future.

Author Contributions: S.X.: Contributed to conceptualization, writing—review and editing, funding acquisition, and supervision. J.L.: Responsible for methodology, software, data curation, and writing—original draft. S.L.: Involved in writing—review and editing. S.Y.: Contributed to investigation, writing—review and editing. F.L.: Assisted with review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Beijing Association for Science and Technology Young Elite Scientist Sponsorship Program (BYESS2023008), the Key Laboratory of Urban Spatial Informatics, Ministry of Natural Resources of the People's Republic of China (2023ZD002), China Scholarship Council (03998521001), and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-05950).

Data Availability Statement: Data availability is not applicable to this article as no new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bates, P.D. Flood inundation prediction. *Annu. Rev. Fluid Mech.* **2022**, *54*, 287–315. [CrossRef]
2. Huang, D.; Wang, S.; Liu, Z. A systematic review of prediction methods for emergency management. *Int. J. Disaster Risk Reduct.* **2021**, *62*, 102412. [CrossRef]
3. Neu, D.A.; Lahann, J.; Fettke, P. A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artif. Intell. Rev.* **2022**, *55*, 801–827. [CrossRef]
4. Zhao, L. Event prediction in the big data era: A systematic survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–37. [CrossRef]

5. Shyalika, C.; Wickramarachchi, R.; Sheth, A. A Comprehensive Survey on Rare Event Prediction. *arXiv* **2023**, arXiv:2309.11356.
6. Alcántara Francia, O.A.; Nunez-del-Prado, M.; Alatrista-Salas, H. Survey of text mining techniques applied to Judicial decisions prediction. *Appl. Sci.* **2022**, *12*, 10200. [CrossRef]
7. Kashpruk, N.; Piskor-Ignatowicz, C.; Baranowski, J. Time Series Prediction in Industry 4.0: A Comprehensive Review and Prospects for Future Advancements. *Appl. Sci.* **2023**, *13*, 12374. [CrossRef]
8. Wallin, J.A. Bibliometric methods: Pitfalls and possibilities. *Basic Clin. Pharmacol. Toxicol.* **2005**, *97*, 261–275. [CrossRef]
9. Chen, C. The citespace manual. *Coll. Comput. Inform.* **2014**, *1*, 1–84.
10. Muthiah, S.; Butler, P.; Khandpur, R.P.; Saraf, P.; Self, N.; Rozovskaya, A.; Zhao, L.; Cadena, J.; Lu, C.-T.; Vullikanti, A. EMBERS at 4 years: Experiences operating an open source indicators forecasting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 205–214.
11. Muthiah, S. Forecasting Protests by Detecting Future Time Mentions in News and Social Media. Ph.D. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 2014.
12. Muthiah, S.; Huang, B.; Arredondo, J.; Mares, D.; Getoor, L.; Katz, G.; Ramakrishnan, N. Planned protest modeling in news and social media. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; pp. 3920–3927.
13. Ramakrishnan, N.; Butler, P.; Muthiah, S.; Self, N.; Khandpur, R.; Saraf, P.; Wang, W.; Cadena, J.; Vullikanti, A.; Korkmaz, G. 'Beating the news' with EMBERS: Forecasting civil unrest using open source indicators. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 1799–1808.
14. Basnet, S.; Soh, L.-K.; Samal, A.; Joshi, D. Analysis of multifactorial social unrest events with spatio-temporal k-dimensional tree-based dbscan. In Proceedings of the 2nd ACM SIGSPATIAL Workshop on Analytics for Local Events and News, Seattle, WA, USA, 6 November 2018; pp. 1–10.
15. Iyda, J.J.; Geetha, P. An improved deep belief neural network based civil unrest event forecasting in twitter. *Appl. Intell.* **2023**, *53*, 5714–5731. [CrossRef]
16. Timoneda, J.C.; Wibbels, E. Spikes and variance: Using Google Trends to detect and forecast protests. *Political Anal.* **2022**, *30*, 1–18. [CrossRef]
17. Lawson, F.H. Repertoires of contention in contemporary Bahrain. In *Islamic Activism: A Social Movement Theory Approach*; Indiana University Press: Bloomington, IN, USA, 2004; pp. 89–111.
18. Laxman, S.; Tankasali, V.; White, R.W. Stream prediction using a generative model based on frequent episodes in event sequences. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2008; pp. 453–461.
19. Rong, H.; Teixeira, A.; Soares, C.G. Maritime traffic probabilistic prediction based on ship motion pattern extraction. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 108061. [CrossRef]
20. Laxman, S.; Sastry, P.; Unnikrishnan, K. Discovering frequent generalized episodes when events persist for different durations. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1188–1201. [CrossRef]
21. Zhou, C.; Cule, B.; Goethals, B. A pattern based predictor for event streams. *Expert Syst. Appl.* **2015**, *42*, 9294–9306. [CrossRef]
22. Yonamine, J.E. Predicting Future Levels of Violence in Afghanistan Districts Using Gdelt. *Unpubl. Manuscr.* 2013. Available online: http://data.gdeltproject.org/documentation/Predicting-Future-Levels-of-Violence-in-Afghanistan-Districts-using-GDELT.pdf (accessed on 15 December 2023).
23. Petroff, V.B.; Bond, J.H.; Bond, D.H.; Bond, D.H. Using hidden Markov models to predict terror before it hits (again). In *Handbook of Computational Approaches to Counterterrorism*; Springer: New York, NY, USA, 2013; pp. 163–180.
24. Salfner, F.; Lenk, M.; Malek, M. A survey of online failure prediction methods. *ACM Comput. Surv. (CSUR)* **2010**, *42*, 1–42. [CrossRef]
25. Carbo-Bustinza, N.; Iftikhar, H.; Belmonte, M.; Cabello-Torres, R.J.; De La Cruz, A.R.H.; López-Gonzales, J.L. Short-Term Forecasting of Ozone Concentration in Metropolitan Lima Using Hybrid Combinations of Time Series Models. *Appl. Sci.* **2023**, *13*, 10514. [CrossRef]
26. Ballı, S. Data analysis of COVID-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos Solitons Fractals* **2021**, *142*, 110512. [CrossRef]
27. Chen, F.; Zhou, B.; Alim, A.; Zhao, L. A generic framework for interesting subspace cluster detection in multi-attributed networks. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; pp. 41–50.
28. Chen, M.; Yu, X.; Liu, Y. PCNN: Deep convolutional networks for short-term traffic congestion prediction. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 3550–3559. [CrossRef]
29. Cheon, S.-P.; Kim, S.; Lee, S.-Y.; Lee, C.-B. Bayesian networks based rare event prediction with sensor data. *Knowl.-Based Syst.* **2009**, *22*, 336–343. [CrossRef]
30. Cho, C.-W.; Zheng, Y.; Wu, Y.-H.; Chen, A.L. A tree-based approach for event prediction using episode rules over event streams. In Proceedings of the Database and Expert Systems Applications: 19th International Conference, DEXA 2008, Turin, Italy, 1–5 September 2008; Proceedings 19. pp. 225–240.

31. Choi, Y.; Choi, M.; Kim, M.; Ha, J.-W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 8789–8797.
32. Korkmaz, G.; Cadena, J.; Kuhlman, C.J.; Marathe, A.; Vullikanti, A.; Ramakrishnan, N. Combining heterogeneous data sources for civil unrest forecasting. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France, 25–28August 2015; pp. 258–265.
33. Zhao, L.; Gkountouna, O.; Pfoser, D. Spatial auto-regressive dependency interpretable learning based on spatial topological constraints. *ACM Trans. Spat. Algorithms Syst. (TSAS)* **2019**, *5*, 1–28. [CrossRef]
34. Tama, B.A.; Comuzzi, M. An empirical comparison of classification techniques for next event prediction using business process event logs. *Expert Syst. Appl.* **2019**, *129*, 233–245. [CrossRef]
35. Xu, S.; Li, S.; Huang, W. A spatial-temporal-semantic approach for detecting local events using geo-social media data. *Trans. GIS* **2020**, *24*, 142–173. [CrossRef]
36. Kattan, A.; Fatima, S.; Arif, M. Time-series event-based prediction: An unsupervised learning framework based on genetic programming. *Inf. Sci.* **2015**, *301*, 99–123. [CrossRef]
37. Chen, F.; Neill, D.B. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 1166–1175.
38. Xu, S.; Li, S.; Huang, W.; Wen, R. Detecting spatiotemporal traffic events using geosocial media data. *Comput. Environ. Urban Syst.* **2022**, *94*, 101797. [CrossRef]
39. Duan, H.; Sun, Z.; Dong, W.; He, K.; Huang, Z. On clinical event prediction in patient treatment trajectory using longitudinal electronic health records. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 2053–2063. [CrossRef] [PubMed]
40. Eria, K.; Marikannan, B.P. Systematic review of customer churn prediction in the telecom sector. *J. Appl. Technol. Innov.* **2018**, *2*, 7–14.
41. Fellbaum, C. WordNet. In *Theory and Applications of Ontology: Computer Applications*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 231–243.
42. Field, C.B. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2012.
43. Flouris, I.; Giatrakos, N.; Deligiannakis, A.; Garofalakis, M.; Kamp, M.; Mock, M. Issues in complex event processing: Status and prospects in the big data era. *J. Syst. Softw.* **2017**, *127*, 217–236. [CrossRef]
44. Chimmula, V.K.R.; Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* **2020**, *135*, 109864. [CrossRef]
45. Kapoor, A.; Ben, X.; Liu, L.; Perozzi, B.; Barnes, M.; Blais, M.; O'Banion, S. Examining COVID-19 forecasting using spatio-temporal graph neural networks. *arXiv* **2020**, arXiv:2007.03113.
46. Deng, S.; Rangwala, H.; Ning, Y. Dynamic knowledge graph based multi-event forecasting. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, 23–27 August 2020; pp. 1585–1595.
47. Kang, W.; Chen, J.; Li, J.; Liu, J.; Liu, L.; Osborne, G.; Lothian, N.; Cooper, B.; Moschou, T.; Neale, G. Carbon: Forecasting civil unrest events by monitoring news and social media. In Proceedings of the Advanced Data Mining and Applications: 13th International Conference, ADMA 2017, Singapore, 5–6 November 2017; Proceedings 13. pp. 859–865.
48. Wang, D.; Ding, W.; Yu, K.; Wu, X.; Chen, P.; Small, D.L.; Islam, S. Towards long-lead forecasting of extreme flood events: A data mining framework for precipitation cluster precursors identification. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1285–1293.
49. Mirtaheri, M.; Abu-El-Haija, S.; Hossain, T.; Morstatter, F.; Galstyan, A. Tensor-based Method for Temporal Geopolitical Event Forecasting. In Proceedings of the ICML Workshop on Learning and Reasoning with Graph-Structured Representations, Long Beach, CA, USA, 9–15 June 2019.
50. Rostami, M.; Huber, D.; Lu, T.-C. A crowdsourcing triage algorithm for geopolitical event forecasting. In Proceedings of the 12th ACM Conference on Recommender Systems, New York, NY, USA, 6 October 2018; pp. 377–381.
51. Li, E.Y.; Tung, C.-Y.; Chang, S.-H. The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system. *Int. J. Med. Inform.* **2016**, *92*, 35–43. [CrossRef]
52. Taleb, N.N. *The Black Swan: The Impact of the Highly Improbable*; Random House: New York, NY, USA, 2007; Volume 2.
53. Runde, J. Dissecting the black swan. *Crit. Rev.* **2009**, *21*, 491–505. [CrossRef]
54. Aven, T. On the meaning of a black swan in a risk context. *Saf. Sci.* **2013**, *57*, 44–51. [CrossRef]
55. Flage, R.; Aven, T. Emerging risk–Conceptual definition and a relation to black swan type of events. *Reliab. Eng. Syst. Saf.* **2015**, *144*, 61–67. [CrossRef]
56. Marsh, T.; Pfleiderer, P. "Black Swans" and the financial crisis. *Rev. Pac. Basin Financ. Mark. Policies* **2012**, *15*, 1250008. [CrossRef]
57. Hanes, E.; Machin, S. Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11. *J. Contemp. Crim. Justice* **2014**, *30*, 247–267. [CrossRef]
58. Antipova, T. Coronavirus pandemic as black swan event. In Proceedings of the International Conference on Integrated Science, Online, 1–2 May 2020; pp. 356–366.

59. Clarke, B.J.; Otto, F.E.; Jones, R.G. Inventories of extreme weather events and impacts: Implications for loss and damage from and adaptation to climate extremes. *Clim. Risk Manag.* **2021**, *32*, 100285. [CrossRef]
60. Hajikazemi, S.; Ekambaram, A.; Andersen, B.; Zidane, Y.J. The Black Swan–Knowing the unknown in projects. *Procedia-Soc. Behav. Sci.* **2016**, *226*, 184–192. [CrossRef]
61. Bedi, J.; Toshniwal, D. CitEnergy: A BERT based model to analyse Citizens' Energy-Tweets. *Sustain. Cities Soc.* **2022**, *80*, 103706. [CrossRef]
62. Goldman, S.A. Limitations and strengths of spontaneous reports data. *Clin. Ther.* **1998**, *20*, C40–C44. [CrossRef]
63. Ismail, H.O. Simultaneous Events and the" Once-Only" Effect. *Front. Artif. Intell. Appl.* **2006**, *150*, 143.
64. Romero, S.; Becker, K. A framework for event classification in tweets based on hybrid semantic enrichment. *Expert Syst. Appl.* **2019**, *118*, 522–538. [CrossRef]
65. Kumar, A.; Singh, J.P.; Dwivedi, Y.K.; Rana, N.P. A deep multi-modal neural network for informative Twitter content classification during emergencies. *Ann. Oper. Res.* **2020**, *319*, 791–822. [CrossRef]
66. Li, J.; Chen, C. *Citespace: Scientific Text Mining and Visualization*; Capital University of Economics and Trade Press: Beijing, China, 2016; pp. 117–119. Available online: https://www.scirp.org/reference/referencespapers?referenceid=2349360 (accessed on 15 December 2023).
67. Liu, X.; Zhao, S.; Tan, L.; Tan, Y.; Wang, Y.; Ye, Z.; Hou, C.; Xu, Y.; Liu, S.; Wang, G. Frontier and hot topics in electrochemiluminescence sensing technology based on CiteSpace bibliometric analysis. *Biosens. Bioelectron.* **2022**, *201*, 113932. [CrossRef] [PubMed]
68. Shao, H.; Kim, G.; Li, Q.; Newman, G. Web of science-based green infrastructure: A bibliometric analysis in citespace. *Land* **2021**, *10*, 711. [CrossRef]
69. Wang, W.; Lu, C. Visualization analysis of big data research based on Citespace. *Soft Comput.* **2020**, *24*, 8173–8186. [CrossRef]