Supervised Learning of the Global Risk Network Activation from Media Event Reports

Xiang Niu1 · Gyorgy Korniss2 · Boleslaw K. Szymanski1

Received: 22 July 2019 / Accepted: 2 September 2019 / Published online: 12 October 2019
© Springer Nature Singapore Pte Ltd 2019

Abstract
The World Economic Forum (WEF) publishes annual reports of global risks which have the high impact on the world’s economy. Currently, many researchers analyze the modeling and evolution of risks. However, few studies focus on validation of the global risk networks published by the WEF. In this paper, we first create a risk knowledge graph from the annotated risk events crawled from the Wikipedia. Then, we compare the relational dependencies of risks in the WEF and Wikipedia networks and find that they share over 50% of their edges. Moreover, the edges unique to each network signify the different perspectives of the experts and the public on global risks. To reduce the cost of manual annotation of events triggering risk activation, we build an auto-detection tool which filters out over 80% media reported events unrelated to the global risks. In the process of filtering, our tool also continuously learns keywords relevant to global risks from the event sentences. Using locations of events extracted from the risk knowledge graph, we find characteristics of geographical distributions of the categories of global risks.

Keywords Global risk network · Knowledge graph · Auto-detection · Supervised learning

Introduction
The World Economic Forum publishes annually the list of global risks and their properties. The global risks are clustered into five categories: economic, environmental, geopolitical, societal, and technological [1]. These risks affect many aspects of lives of countless people. Economic risks may inflict tremendous damage on individuals, financial institutions, and governments. For example, in 2008, the asset bubble in several countries caused most of the major economies to undergo a recession lasting several years [2]. Environmental risks affect both individuals and infrastructure but their impact is global when thousands of people die or are displaced by hurricanes or earthquakes [3]. Geopolitical risks are mainly caused by a failure of national or global governance but may result in millions of people suffering from war [4], hunger [5], and poverty [6]. Technological risks are real and often unpredictable consequences of advanced technologies such as cyberattacks [7] and data leaks [8] threatening well-being of individuals and corporations. In short, all risks have critical ability to disrupt civilization at diverse levels, ranging from individuals to nations, so there is an urgent need to develop tools for their discovery, analysis, and prediction.

The big challenge to progress in this direction is to identify activation of global risks from massive event records in Web media. Therefore, risk-event labeling is a fundamental and necessary preparatory step before any in-depth study of risks, such as risk modeling [9] and control [10], can be attempted. A high-quality dataset combined with the validated models is a foundation for global risks reliable discovery, analysis, and prediction. The first systematic approach
for such data collection effort was done and described in [11], in which the researchers manually collected monthly data points for each risk using online event resources such as articles, Wikipedia and news to annotate events relevant to any global risk activation. The covered period ranged from the beginning of year 2000 to the end of year 2016. Considering the inaccuracy and incompleteness of manually labeled risk event records, researchers further improve these aspects of collected data by providing a detailed reason for selecting each risk event [12, 13]. However currently, manual annotation is the only way to label risk events because of the difficulty to automate this operation caused by the unstructured format of event description. Here, we first propose a human–machine interaction model to augment the manual annotation process by precisely filtering out a large number of unrelated event. Then, we enhance this process by continuously expanding keyword sets for events using the supervised online learning.

The auto-detection tool helps analysts to detect every single risk from event descriptions in Web media. However, individual risk analysis is not sufficient because rarely a risk activation happens in isolation. According to the annual report from the World Economic Forum, the connectivity among risks is common and complicated. The global risks are clustered into five categories but their links include both intra-category and inter-categories edges, both types of which are common. Most of the societal risks can activate under influence of economic (profound political and social instability), environmental (rapid and massive spread of infectious diseases) or geopolitical (large-scale involuntary migration) risks. The yearly WEF Global Risk Report is based on the experts’ answers to the annual questionnaires that ask about each risk probability of activation, such activation costs to global economy and interconnectivity between risks. This dependence on expert opinions raises one important question: “how well do the experts estimate risk probability of activation, impact, and interconnectivity?” followed by another: “are their networks reliable, integrated, and representative?” To partially answer these questions, here, we create another risk network based only on the risk events discovered in Web media data. Then, we compare the resulting network with the WEF global risk network for the same risk nodes to assess both networks correctness and authoritative. We expect that the network from event data will be biased towards risks with high level of public awareness, which tend to be directly related to human lives, such as “Asset bubble” and “Extreme weather event”.

Although the global risks attempt to represent events which have direct impact on global economy and systematic impact on the entire world, most of the time, the risk events occur locally and have regional effects and influence. Thanks to the specific characteristic of the risk-events dataset, we extract local information from each event record and build connections between the locations and related risks. Having collected data for the period from year 2000 to 2014, we illustrate the different regional patterns of five risk categories on the world political maps.

### Risk Detection

In [11], the authors investigate multiple event resources including articles, Wikipedia, and news. Here, we only focus on the Wikipedia Current Events Portal (WCEP) [14], which comprehensively includes daily summaries of news events edited by crowdsourcing. Because of its popularity and completeness, the WCEP has been fully analyzed by comparing it with content written by professional journalists in WikiTimes project [15]. The conclusion of this project is that the WCEP is a representative source of public news thanks to its stable contribution volumes and crowdsourcing. The WikiTimes [15] also provides specific crawling methods and extracted dataset with over 50 k events from the WCEP from the year 2000–2014, which are sufficient enough for the risk-event labeling and further studies. In the dataset, each event contains information detailed in Table 1. Figure 1 shows sample subnetwork of events. The event 4359, with description “Tropical Storm Gaston douses Richmond, Virginia, with up to 14 inches of rain, causing widespread flooding” and entities “Tropical Storm Gaston”, “Virginia” and “Mark Warner”, is part of story “2004 Atlantic hurricane season” occurring on Aug. 30, 2004, reported by ABC-NEWS. The event 4571 is described as “At least nine deaths in Florida, two deaths in the Bahamas, and one death in Georgia are blamed on the storm. Damage estimates range widely from US$2 to US$15 billion”. The event 4622 reports “The Cuban government evacuates between 800,000 and 1.3 million people from coastal cities and developed areas”. The last two events are parts of the same story about “2004 Atlantic hurricane season”.

Manual annotation requires a person to go first through the event description word by word and select the risk related keywords, then label the event with risk tags according to the keywords. For example, for the event 4359, two keywords “storm” and “flooding” may be selected and then

| Concept           | Description                                      |
|-------------------|--------------------------------------------------|
| Event             | Sentences of a news summary                      |
| Story             | A story in Wikipedia of which the event is part   |
| Category          | A category to which the event belongs             |
| Date              | The event date of occurrence                     |
| Entities          | A list of entities in the event sentences         |
| References        | External sources for the same event              |
the event will be labeled with tags “storm” and “flood” accordingly. Both tags belong to the same risk “10. Extreme weather events”, as shown Table 5 in Appendix. The event 4571 contains two more tags “death” and “economic loss” in addition to “storm”. Thus, this event shows the activity of one more risk “11. Failure of climate-change mitigation and adaptation”, see Table 5 in Appendix. The event 4622 contains only one tag “evacuation”, which belongs to risk “23. Large-scale involuntary migration”, as shown in Table 7 in Appendix. Thus, the three events represent the activities of three risks in two risk categories: environmental risks, as seen in Table 5 in Appendix, and societal risks according to Table 7 in Appendix.

With the development of modern information society, there are a large number of news events described all over the internet. Specifically, around three thousand new events are generated in the WCEP every year amounting to about 250 per each month and over fifty thousand events in the past 15 years. Given 29 risk labels and around 20 words per event description, an expert would have to go through $3000 \times 30 \times 20$ so nearly two million comparisons to process just one year of events data in the WCEP, which is a heavy workload. Besides, during the labeling, only a few (around two thousand) of the fifty thousand events are relevant to global risks. These observations motivated us to create an auto-detection tool to help analysts by filtering out the irrelevant events, thereby saving time in risk-event labeling. Our risk auto-detection module is depicted in Fig. 2. The detailed processing proceeds according to the following steps:

Fig. 1 An example of risk knowledge graph consisting of three events. Event 4359: “Tropical Storm Gaston douses Richmond, Virginia, with up to 14 inches of rain, causing widespread flooding (8/30/04)”. Event 4571: “At least nine deaths in Florida, two deaths in the Bahamas, and one death in Georgia are blamed on the storm. Damage estimates range widely from US$2 to US$15 billion (9/7/04)”. Event 4622: “The Cuban government evacuates between 800,000 and 1.3 million people from coastal cities and developed areas (9/13/04)”
Step 1 Initially, an analyst prepares a few keywords and tags for each risk according to the description of 29 risks in Appendix based on [1].

Step 2 If the given list of new pairs with keyword \( k \) and tag \( t \) \( \langle k, t \rangle \) is empty, processing terminates, otherwise, these pairs are merged into keyword dictionary.

Step 3 For each event \( e \) and a list of keywords \( K \), each paired with some tag \( t \), if a keyword \( k \in K \) appears in the event sentence manually accepted as relevant, the event \( e \) is labeled with the tag \( t \) as \( \langle e, t^+ \rangle \), otherwise, the event \( e \) is labeled as \( \langle e, t^- \rangle \).

Step 4 We use a bag-of-words representation \( b = \langle w_1, c_1 \rangle, \ldots, \langle w_n, c_n \rangle \) for each event sentence, where \( w_i \) is the \( i \)th word of event sentence in \( e \), \( c_i \) is the count of appearances of the word \( w_i \) for \( n \) unique words. For each tag \( t \), we have a list of events: \( \langle e, t^+ \rangle \) is a positive instance with bag-of-words representation \( b^+ \in B^+ \), while \( \langle e, t^- \rangle \) is a negative instance with bag-of-words representation \( b^- \in B^- \) in the event sentence.

Step 5 With random forest classifier given training dataset \( B^+ \) and \( B^- \), we get entropy ranking of featured keywords \( w_1, w_2, \ldots, w_m \), where \( m \) is the total number of unique words in \( B^+ \) and \( B^- \). The words with high rank are important for the classification.

Step 6 Based on the entropy ranking list, we select meaningful keywords from top \( a \) (default \( a = 5 \)) new ones and create a list of \( \langle k, t \rangle \). Then, processing restarts at step 2.

Tables 4, 5, 6, 7, 8 in Appendix show the list of selected tags and keywords of 29 risks with current dataset up to year 2014. With new events data, the module can update the list further and improve the filtering accuracy. In step 3, the events without any keyword \( k \in K \) or rejected by analysts are filtered out as non-relevant events.

Risk Network

The auto-detection module produces a structured event dataset containing rich independent information such as risk labels, stories, event descriptions, their occurrence locations, and dates. Based on that, with the specific principal (story) and subordinate (event) relationship, we further build event correlations, see Fig. 1. Any two events are considered related when they belong to the same story. The knowledge graph includes the following four types of relationships:

- At date of: an event \( e \) happened at date \( d \). Any event \( e \) has a unique \( d \) associated with it. But one date \( d \) may correspond to multiple events.
- At location of: an event \( e \) happened at location \( l \) or at location relevant to \( l \). Any event \( e \) may happen (affect) multiple locations, e.g., Fig. 1 event 4571 affected locations at Bahamas, Florida and Georgia. One location \( l \) may also correspond to multiple events.
- Prior to: an event \( e_1 \) happened prior to another event \( e_2 \) when they are in the same story and the date \( d_1 \) of \( e_1 \) precedes the date \( d_2 \) of \( e_2 \).
- In risk of: an event \( e \) is in risk of \( t \) after the event sentence is labeled by the tag \( t \) by the auto-labeling module, see Fig. 2.
Figure 3 depicts the entire knowledge graph that contains all relationships between any two events, and between two pairs for event $e_1, (e_1, \text{risk})$ and $(e_1, \text{date})$. Here, we group all tags $t_1, t_2, \ldots$ belonging to the same risk $r$ into one risk node. For example, in Fig. 1, event 4359 is in risk of “storm” and “flood”, which correspond to the same risk “10. Extreme weather event”.

Additionally, we extracted the risk network (Wiki in Fig. 4) from the entire knowledge graph (Fig. 3). For each pair of events $(e_1, e_2)$, when $e_1$ is prior to $e_2$, $e_1$ is in risks of $R_1$ and $e_2$ in risks of $R_2$, we create an edge between each pair of $(r_i, r_j)$, $r_i, r_j \in R_1 \cup R_2$, $i \neq j$. For example in Fig. 1, event 4359 is in risks of “10. Extreme weather event” $r_{10}$, event 4571 is in risks of “10. Extreme weather event” $r_{10}$ and “11. Failure of climate-change mitigation and adaptation” $r_{11}$, event 4622 is in risks of “23. Large-scale involuntary migration” $r_{23}$. Thus, we have three connections $r_{10} - r_{11}$, $r_{10} - r_{23}$ and $r_{11} - r_{23}$. To make a comparison of so created network with the undirected and unweighted risk network created by the World Economic Forum (WEF) in report [1], we transform the Wiki network to be also undirected and unweighted, as seen in Fig. 4, where the size of each risk node represents its degree. In general, the two networks are similar in nodes sizes and connections. There are 170 edges in WEF network and 224 edges in Wiki network, with the number of common edges as high as 121, see Table 2. Specifically, the risk “23. Profound social instability” has the largest number (18) of common related risks between the two networks.

We also highlight the different risk dependencies in WEF and in Wiki networks in Fig. 4. For example, in Wiki network, risk “1. Asset bubbles in a major economy” influences and is influenced also by risks “4. Failure/shortfall of critical infrastructure”, “22. Large-scale involuntary migration”, etc.
The degree differences between the two risk networks arise from the different views of the risks by experts and the public. The risks with a higher degree in Wiki network tend to be a focus of public concerns, such as “1. Asset bubbles in a major economy”, “10. Extreme weather events”, “17. Large-scale terrorist attacks”, “21. Food crises” and “29. Massive incident of data fraud/theft”. Nevertheless, the risks with a higher degree in WEF network tend to be a focus of expert concerns, often novel or merely newly arising, such as “2. Deflation in a major economy”, “12. Major biodiversity loss and ecosystem collapse”, and “26. Adverse consequences of technological advances”. In addition, in Table 2, the common part of the two networks includes most of the intra-group edges. After extracting it from the WEF and the Wiki network, almost no intra-group edges are left, especially in Wiki network. Moreover, in Table 3, after extracting the common part, the WEF network becomes fragmented with extremely low average clustering coefficients in each group and in the entire network.

### Table 2 The number of intra-group and inter-group edges of each risk category and network

| Networks     | Common In | Common Out | WEF only In | WEF only Out | WEF In | WEF Out | Wiki only In | Wiki only Out | Wiki In | Wiki Out |
|--------------|-----------|------------|-------------|--------------|-------|---------|-------------|--------------|--------|---------|
| Economic     | 21        | 36         | 4           | 13           | 25    | 49      | 6           | 50           | 27     | 86      |
| Environmental| 4         | 22         | 6           | 14           | 6     | 14      | 0           | 28           | 4      | 50      |
| Geopolitical | 8         | 29         | 2           | 11           | 10    | 40      | 0           | 34           | 8      | 63      |
| Societal     | 12        | 47         | 2           | 13           | 14    | 60      | 1           | 45           | 13     | 92      |
| Technological| 3         | 12         | 3           | 13           | 6     | 25      | 0           | 35           | 3      | 47      |
| Total edges  | 121       | 48         | 170         | 103          | 224   |         |             |              |        |         |

WEF represents the network published by the World Economic Forum; Wiki represents the network created by us from Wiki events. Common part represents the edges existing in both networks. WEF only represents the edges unique to the WEF network, while Wiki only includes the edges unique to the Wiki network.

### Risk Activities Heat Maps Layered Over the Political Maps of the World

Using the risk knowledge graph, we extract connections between locations and risks. For any event $e$ in risk $r \in R$ that occurred at location $l \in L$, where $R$ is one of the five risk categories, $L$ is a country to which location $l$ belongs, we record an occurrence of pair $(R, L)$. After processing all events, we obtain the number of occurrences of each pair $(R, L)$. For each risk category $R$, we normalize the number of occurrences in all countries $L$ by taking logarithm of a ratio of number of occurrence $+1$ in the given country to $1+$ the maximum number of occurrences in any country for category $R$. We show the results as a heat maps layered over the political maps of the world shown in Figs. 5, 6, 7. Each of these figures displays normalized levels of activities of one of the five risk categories. For example, in economic risks, the economic risk events related to Russia occur 42 times, related to Japan 16 times, to India 6 times, to Italy 2 times. These scores after the normalization become 1, 0.75, 0.52, 0.29, respectively.

As shown in Fig. 5a, most of the economic risks happen in the developed countries or countries with high volume

### Table 3 The average degree $\langle k \rangle$ and average clustering coefficient $\langle c \rangle$ of each risk category and network

| Networks     | Common $\langle k \rangle$ | Common $\langle c \rangle$ | WEF only $\langle k \rangle$ | WEF only $\langle c \rangle$ | WEF $\langle k \rangle$ | WEF $\langle c \rangle$ | Wiki only $\langle k \rangle$ | Wiki only $\langle c \rangle$ | Wiki $\langle k \rangle$ | Wiki $\langle c \rangle$ |
|--------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-------------------------|-------------------------|-----------------------------|---------------------------|-------------------------|-------------------------|
| Economic     | 8.7                         | 0.60                        | 2.3                         | 0.00                        | 11.0                    | 0.61                    | 6.9                         | 0.35                      | 15.6                    | 0.74                    |
| Environmental| 6.0                         | 0.49                        | 5.2                         | 0.06                        | 11.2                    | 0.66                    | 5.6                         | 0.42                      | 11.6                    | 0.81                    |
| Geopolitical | 9.0                         | 0.69                        | 3.0                         | 0.31                        | 12.0                    | 0.57                    | 6.8                         | 0.35                      | 15.8                    | 0.81                    |
| Societal     | 11.8                        | 0.59                        | 2.8                         | 0.17                        | 14.7                    | 0.59                    | 7.8                         | 0.32                      | 19.7                    | 0.72                    |
| Technological| 4.5                         | 0.47                        | 4.8                         | 0.01                        | 9.3                     | 0.48                    | 8.8                         | 0.22                      | 13.3                    | 0.85                    |
| Whole network| 8.3                         | 0.58                        | 3.3                         | 0.10                        | 11.7                    | 0.59                    | 7.1                         | 0.34                      | 15.5                    | 0.78                    |

WEF represents the network published by the World Economic Forum; Wiki represents the network created by us from Wiki events. Common part includes the edges existing in both networks. WEF only represents the edges unique to the WEF network, while Wiki only includes the edges unique to the Wiki network.
of trading resources, industrial products or services, such as Russia (1.00), U.S. (0.95), Malaysia (0.78), Japan (0.75), Mexico (0.72), Cyprus (0.68), Australia (0.61), Greece (0.61). The most impactful story in Russia is “2008 Russian financial crisis”, in the U.S. “Subprime mortgage crisis”, in Malaysia “Malaysia Airlines Flight 370”, in Japan “Automotive industry crisis of 2008–2010”, in Mexico “Mexican Drug War”, in Cyprus “2012–2013 Cypriot financial crisis”, in Australia “August 2011 stock markets fall”, and in Greece “Greek government-debt crisis”. Those economic risks are related to a failure of financial mechanisms such as stock (Russia, Australia), mortgage (U.S.), banking (Cyprus, Greece), illicit trade, such as smuggling (Mexico), and decreased demand for manufacturing products, such as cars (Japan).

Fig. 5  Heat maps of a economic and b environmental risks overlyed on the political map of the world

Fig. 6  Heat maps of a geopolitical and societal b risks overlayed on the political map of the world

Fig. 7  Technological risks heat map over political map of the world
The environmental risks, shown in Fig. 5b, usually happen in coastal countries, especially those near Atlantic, Pacific, or Indian Oceans, such as U.S. (1.00), Philippines (1.00), China (0.97), Japan (0.91), India (0.91), Mexico (0.90), Haiti (0.87), Pakistan (0.83), Cuba (0.71). “Atlantic hurricane season” mainly affects the U.S., Mexico, Haiti, Cuba. “Pacific typhoon season” mainly affects the Philippines, China, Japan. “North Indian Ocean cyclone season” mainly affects India. Beyond that, some other disasters also happen in the U.S. “2011 Mississippi River floods”, in China “2008 Sichuan earthquake”, in Japan “2011 Tōhoku earthquake and tsunami”, in India “2004 Indian Ocean earthquake and tsunami”, in Mexico “Deepwater Horizon oil spill”, and in Pakistan “2010 Pakistan floods”.

As shown in Fig. 6a, the geopolitical risks are most frequent in countries having unstable regime or involved in international conflicts, such as North Korea (1.00), Thailand (0.80), Iran (0.80), U.S. (0.72), Egypt (0.72), Pakistan (0.62), Syria (0.62), Nigeria (0.62), Iraq (0.38). Risk “16. Interstate conflict with regional consequences” is between the U.S. and Iraq in “Iraq War”. Risk “17. Large-scale terrorist attacks” is in Pakistan “Terrorism in Pakistan” and Nigeria “Islamic insurgency in Nigeria”. Risk “18. State collapse or crisis” is in Thailand “Thai coup d'état”, Egypt “Egyptian coup d’état”, and Syria “Syrian civil war”. Risk “19. Weapons of mass destruction” is in North Korea “North Korean nuclear test” and Iran “Nuclear program of Iran”.

Societal risks, shown in Fig. 6b, are often a consequence of geopolitical risks and usually occur in the same countries as the latter group of risks does, such as Egypt (1.00), Syria (0.98), Yemen (0.80), Libya (0.77), Tunisia (0.77), China (0.76), Russia (0.76), Saudi Arabia (0.73). The most influential story in Egypt is “Egyptian Revolution of 2011”, in Syria “Syrian civil war”, in Yemen “Yemeni Revolution”, in Libya “Libyan civil war”, in Tunisia “Tunisian Revolution”, in China “2009 Ürümqi riots”, in Russia “Snow Revolution”, and in Saudi Arabia “Arab Spring”.

As shown in Fig. 7, technological risks frequently happen either in a scientifically developed country such as U.S. (1.00) and U.K. (0.51) or in unstable nations such as Egypt (0.85), Syria (0.66), Iraq (0.51), Pakistan (0.51), North Korea (0.51). When information technology is highly developed, risk “29. Massive incident of data fraud/theft” is often active because of a massive amount of official and personal data online, such as “United States diplomatic cables leak” in the U.S. and “News International phone hacking scandal” in the U.K. Related geopolitical risks may also cause other technological risks. For example, in Egypt, a long-lasting internet blackout and censorship was triggered by “Egyptian Revolution of 2011”; in North Korea, risk “28. Large-scale cyberattacks” was caused by “2013 Korean crisis”.

Conclusion

Our paper contributes the first event-driven analysis of global risks using the complete Wikipedia Event Portal dataset over the years 2000–2014. To overcome the limitation of risk-event manual annotation used in the current research [11, 13], here we introduce a novel risk auto-detection module, which filters out over 80% of data about the non-relevant events and reduces significantly volume of data to be manually labeled. This approach can be used in other machine learning scenarios when the true-negative instances account for a large proportion of the dataset during the supervised learning classification. With the help of auto-detection tool and sufficient effort devoted to manual labeling, we created the complete and structured dataset from Wikipedia Event Portal, which can be studied in depth in the future. This data set contains two thousand labeled risk events extracted from over 50 thousand events. These labeled risks are structured hierarchically with events linked to stories, locations, dates, categories, and references creating a knowledge network.

From this rich dataset, we build a relational network, which is surprisingly similar to the network created by the WEF experts [1]. Moreover, the differences between the two networks reflect the differences in risk views of the expert and the public. The experts show concerns about yet unseen risks that may arise in the future or have an enormous impact on all humanity, while the public tends to focus on risks that matter more to everyday life. By studying the local effects of risks, we find that the economic risks arise in countries with high volume of trade. The environmental risks are common in coastal areas. The geopolitical risks are concentrated mainly in the Middle East. Most of the societal risks are the consequences of geopolitical risks and thus occur in the same countries as the former. The technological risks arise in scientifically developed countries and in the countries in conflict with other nations.

Acknowledgements

This work was supported in part by the Army Research Office (ARO) under Grant W911NF-16-1-0524, by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 (NS CTA), and by the Office of Naval Research (ONR) under Grant N00014-15-1-2640. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied of the Army Research Laboratory or the U.S. Government. Appendix

Tables 4, 5, 6, 7, 8 contain subordinated tags for each risk within one of the five categories of risks.
### Table 4  Economic risks and subordinate tags

| ID | Risk                                         | Tag                      | Keyword                                |
|----|----------------------------------------------|--------------------------|----------------------------------------|
| 1  | Asset bubbles in a major economy             | Housing bubble           | Housing, mortgage, estate              |
|    |                                              | Stock bubble             | Share, stock, exchange                 |
|    |                                              | Commodity bubble         | Commodity-                             |
|    |                                              | Asset bubble             | Liquidity, asset                       |
| 2  | Deflation in a major economy                 | Deflation                | Deflation                              |
| 3  | Failure of a major financial mechanism or institution | Decline in prices | Price                                  |
|    |                                              | Bank failure             | Bank                                   |
|    |                                              | Financial crisis         | Financial                               |
| 4  | Failure/shortfall of critical infrastructure  | Oil/gas network failure  | Pipe                                   |
|    |                                              | Water supply failure     | Pipe                                   |
|    |                                              | Power grid failure       | Power, grid, electricity, blackout     |
|    |                                              | Civil flight failure     | Flight, airport, airline, plane, aircraft |
| 5  | Fiscal crises in key economies               | Sovereign debt crisis    | Debt, loan, lend                        |
|    |                                              | Liquidity crisis         | Liquid                                 |
| 6  | High structural unemployment or underemployment | Unemployment            | Employ, job                            |
| 7  | Illicit trade                                | Drug trade               | Drug                                   |
|    |                                              | Smuggling                | Smuggl-                                |
|    |                                              | Tax evasion               | Tax                                    |
|    |                                              | Money laundering         | Launder                                |
|    |                                              | Human trafficking        | Trafficking                            |
|    |                                              | Counterfeiting           | Counterfeit                            |
| 8  | Severe energy price shock                    | Energy price shock       | Price, oil, gas                        |
| 9  | Unmanageable inflation                       | Inflation                | Inflation                              |
|    |                                              | Rising good price        | Price                                  |

Each tag corresponds to a list of the roots of keywords

### Table 5  Environmental risks and their corresponding tags

| ID | Risk                                         | Tag                      | Keyword                                |
|----|----------------------------------------------|--------------------------|----------------------------------------|
| 10 | Extreme weather events                       | Storm                    | Hurricane, storm, typhoon, cyclone     |
|    |                                              | Blizzard                 | Snow, blizzard, hailstone              |
|    |                                              | Torrential rain          | Rain                                   |
|    |                                              | Flood                    | Flood                                  |
| 11 | Failure of climate-change mitigation and adaptation | Death              | Death, dead, die, toll, kill, corpse   |
|    |                                              | Missing                  | Missing, disappear                     |
|    |                                              | Injury                   | Injur-, hurt, wound                     |
|    |                                              | Homeless                 | Home                                   |
|    |                                              | Damage                   | Collaps, damage, destroy               |
|    |                                              | Economic loss            | Cost, loss, dollar, usd, euro, $, €     |
| 12 | Major biodiversity loss and ecosystem collapse | Osytem collapse          | Ecosystem                               |
|    |                                              | Biodiversity loss        | Biodiversity, species, wildlife        |
| 13 | Major natural disasters                      | Earthquake               | Earthquake, temblor                    |
|    |                                              | Volcanic activity        | Volcan-                                 |
|    |                                              | Tsunami                  | Tsunami                                |
|    |                                              | Landslide                | Slide                                  |
|    |                                              | Natural wildfire         | Wildfire                               |
| 14 | Man-made environmental damage and disasters  | Radioactive contamination | Radioact-                              |
|    |                                              | Environment contamination | Contamin-                              |
|    |                                              | Urban pollution          | Pollution                               |
|    |                                              | Oil spill                | Oil, spill                             |
|    |                                              | Manmade wildfire         | Wildfire                               |

Each tag must match at least one entry in the list of the roots of keywords
### Table 6  Geopolitical risks and their corresponding tags

| ID | Risk                                    | Tag             | Keyword                        |
|----|-----------------------------------------|-----------------|--------------------------------|
| 15 | Failure of national governance          | Corruption      | Corruption                     |
| 16 | Interstate conflict with regional conse-| Trade war       | Trade, currency                |
|    | quences                                 | Military war    | Military, invasion, occupation, conflict |
| 17 | Large-scale terrorist attacks            | Terror          | Terror                         |
| 18 | State collapse or crisis                | Military coup   | Coup                           |
|    |                                         | Civil war       | Civil, war                     |
| 19 | Weapons of mass destruction             | Nuclear weapon  | Nuclear                        |

Each tag must match at least one entry in the list of the roots of keywords.

### Table 7  Societal risks and their corresponding tags

| ID | Risk                                    | Tag             | Keyword                        |
|----|-----------------------------------------|-----------------|--------------------------------|
| 20 | Failure of urban planning               | Urban pollution | Pollution                      |
|    |                                         | Factory explosion | Factory                      |
|    |                                         | Building collapse | Collapse                      |
| 21 | Food crises                             | Starvation      | Starv-, hunger, famin-         |
|    |                                         | Crop destruction | Crop, agricult-               |
|    |                                         | Low quality of food | Poison, taint, contaminat-, poultry, meat, pork |
|    |                                         | Rising food price | Price                         |
|    |                                         | Food shortage    | Shortage, suppl-              |
| 22 | Large-scale involuntary migration        | Evacuation      | Evacuat-, evacuee             |
|    |                                         | Refugee flee    | Refugee, flee                 |
|    |                                         | Migration       | Migrant, migrate              |
|    |                                         | Eviction        | Evict                         |
| 23 | Profound social instability             | Riot            | Riot                           |
|    |                                         | Unrest          | Unrest                         |
|    |                                         | Demonstration   | Demonstration                 |
|    |                                         | March           | March                          |
|    |                                         | Revolution      | Revolution                     |
| 24 | Rapid and massive spread of infectious  | Avian influenza | Bird, avian, h5n1, h6n1, h7n9 |
|    | diseases                                | Swine influenza | h1n1, swine                   |
|    |                                         | Equine influenza | Equine                        |
|    |                                         | Aids            | Aids, hiv                      |
|    |                                         | Cholera         | Cholera                        |
|    |                                         | Fever           | Fever                          |
|    |                                         | Diarrhea        | Diarrhea                       |
|    |                                         | Ebola           | Ebola                          |
|    |                                         | Epidemic        | Epidemic, outbreak, pandemic  |
| 25 | Water crises                            | Water shortage  | Thirst, shortage, deprivat-, suppl-|
|    |                                         | Water contamination | Water                       |

Each tag must match at least one entry in the list of the roots of keywords.
References

1. World Economic Forum Global Risks Report. https://www.weforum.org/reports/the-global-risks-report-2016 (2016). Accessed 31 July 2018
2. Erkens DH, Hung M, Matos P. Corporate governance in the 2007–2008 financial crisis: evidence from financial institutions worldwide. J Corp Finance. 2012;18(2):389.
3. Pielke RA Jr, Gratz J, Landsea CW, Collins D, Saunders MA, Musulin R. Normalized hurricane damage in the United States: 1900–2005. Nat Hazards Rev. 2008;9(1):29.
4. Lee Ray J. Explaining interstate conflict and war: what should be controlled for? Confl Manag Peace Sci. 2003;20(2):1.
5. Sanchez PA. Soil fertility and hunger in Africa. Science. 2002;295(5562):29.
6. Pogge T. World poverty and human rights. Eth Int Aff. 2005;19(1):1.
7. Rid T, Buchanan B. Attributing cyber attacks. J Strateg Stud. 2015;38(1–2):4.
8. Harding L. What are the Panama papers? A guide to history’s biggest data leak. Guardian. 2016. p. 5. https://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers. Accessed 10 Oct 2019
9. Lin X, Moussawi A, Korniss G, Bakdash IZ, Szymanski BK. Limits of risk predictability in a cascading alternating renewal process model. Sci Rep. 2017;7(1):6699.
10. Baloi D, Price AD. Modelling global risk factors affecting construction cost performance. Int J Proj Manag. 2003;21(4):261.
11. Szymanski BK, Lin X, Asztalos A, Sreenivasan S. Failure dynamics of the global risk network. Sci Rep. 2015;5:10998.
12. Niu X, Moussawi A, Derzsy N, Lin X, Korniss G, Szymanski BK. Evolution of the global risk network mean-field stability point. In: Cherifi C, Cherifi H, Karsai M, Musolesi M, editors. International conference on complex networks and their applications. Cham: Springer; 2018. p. 1124–34.
13. Niu X, Moussawi A, Korniss G, Szymanski BK. Evolution of threats in the global risk network. Appl Netw Sci. 2018;3(1):24.
14. Wikipedia event portal. https://en.wikipedia.org/wiki/Portal:Current_events (2019). Accessed 22 May 2019
15. Tran GB, Alrifai M. Indexing and analyzing wikipedia’s current events portal, the daily news summaries by the crowd. In: Proceedings of the 23rd international conference on world wide web. New York: ACM; 2014. p. 511–6

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.