Assessing systemic risk in financial markets using dynamic topic networks

Mike K. P. So1, Anson S. W. Mak2 & Amanda M. Y. Chu3

Systemic risk in financial markets refers to the breakdown of a financial system due to global events, catastrophes, or extreme incidents, leading to huge financial instability and losses. This study proposes a dynamic topic network (DTN) approach that combines topic modelling and network analysis to assess systemic risk in financial markets. We make use of Latent Dirichlet Allocation (LDA) to semantically analyse news articles, and the extracted topics then serve as input to construct topic similarity networks over time. Our results indicate how connected the topics are so that we can correlate any abnormal behaviours with volatility in the financial markets. With the 2015–2016 stock market selloff and COVID-19 as use cases, our results also suggest that the proposed DTN approach can provide an indication of (a) abnormal movement in the Dow Jones Industrial Average and (b) when the market would gradually begin to recover from such an event. From a practical risk management point of view, this analysis can be carried out on a daily basis when new data come in so that we can make use of the calculated metrics to predict real-time systemic risk in financial markets.

Systemic risk is usually associated with the breakdown of systems due to global events, catastrophes, or extreme incidents. In financial markets, systemic risk can be evidenced by the simultaneous declines in the prices of most or all of the entities in the system due to a large-scale collapse, as opposed to the breakdowns of a few entities1–3. As businesses and financial institutions nowadays operate in a complex and dynamic world, their interconnectedness greatly facilitates the propagation of systemic risk. In the 2015–2016 stock market selloff4,5 and in the face of COVID-19, countries have shown a significant increase in systemic risk6,7. There are various ways to measure systemic risk; among these measures is the Chicago Board Options Exchange’s CBOE Volatility Index, also widely known as VIX8, which is designed to capture the level of risk and fear in the market and to generate 30-day-ahead predictions of market volatility9,10. Research has previously shown that VIX indicated higher uncertainty and risk aversion in financial markets during several financial crises in the past, such as the Asian financial crisis in 1997 and the collapse of Lehman Brothers in 200811. However, despite the wide application of and research on VIX, some scholars have argued that the index might underestimate volatility12. Given that VIX is derived from the S&P 500 index, it has also been argued that VIX does not cause the realised volatility in the index: in other words, it is not forward looking13. Systemic risk has been studied using various approaches, including network analysis14–19, spatial-temporal methods20, machine-learning methods21, and changes in cross-correlations22. In this paper, we propose a dynamic topic network (DTN) approach through the combination of two techniques—topic modelling and network analysis—to assess systemic risk in financial markets and also to provide an indication of when the market would slowly recover after the occurrence of a significant event.

The 2015–2016 stock market selloff was partly attributed to the market’s worry on the slowdown of the world economy. The selloff was also partly triggered by the end of the quantitative easing in the United States (U.S.) in 2014, the European sovereign debt crisis and the drop in commodity prices, causing substantial adjustment and turbulence in the stock markets in 2015–2016. Coronavirus disease 2019 (COVID-19) was declared a global pandemic by the World Health Organization (WHO) on 11 March 2020. By the end of October 2021, there had been more than 240 million cases of COVID-19 and 4.8 million deaths caused by the virus, with the numbers expected to keep increasing23. The world has been impacted significantly by the pandemic, and the financial markets are among those whose stability has been undermined. A significant number of studies have been conducted, all of which indicate that COVID-19 has brought volatility to different financial markets across the globe24–27. The 2015–2016 stock market selloff and COVID-19 are examples of systemic risk, which in finance

1Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology, Hong Kong, China. Faculty of Science, University of Amsterdam, 1098XH Amsterdam, The Netherlands. Department of Social Sciences, The Education University of Hong Kong, Hong Kong, China. Email: immkpso@ust.hk
refers to the risk that a shock or failure event triggers the breakdown of the entire system; the scale could be
domestic or transnational. We are interested in applying our DTN method to study the systemic risk during the
2015–2016 stock market selloff and the COVID-19 pandemic.

Research has been conducted to extract insights from textual information: for instance, to explain abnormal
volatility in financial markets through news18, to forecast market stress and volatility19–20, to improve the accuracy of
predictions in time series volatility models31, and to correlate Internet searches with subsequent movement in
stock prices42. Sentiment analysis is another way to make use of textual information: for instance, the use of
Twitter feeds to predict the Dow Jones Industrial Average (DJI A)35 and the use of posts on Weibo to predict
Chinese stock price movement34. These studies suggested that textual information provides certain predictive
powers in financial markets. On the other hand, some studies have used network analysis to depict relations
between different entities35. Initially, network analysis was applied predominantly in the field of social sciences36,
but it has now been extended to other fields of study: for example, in the field of medicine to understand the
transmission of infectious diseases37, the relations between genetics and human diseases38, and drug-target
interactions39 in the finance field to assess systemic risk and contagion effects in financial markets40. Due
to the global pandemic, an increasing number of studies have also made use of network analysis to predict and
visualise pandemic risk41–43 and its influence on financial market connectedness44–46.

Although real-time financial news is readily available, research combining textual information and network
analysis to predict financial markets is limited. Some similar approaches have previously been proposed. One
study, for example, made use of the topic modelling technique to evaluate the temporal change in extracted
topics using Japanese news46; in that paper, however, articles are grouped by month before the analysis, so the
networks might not be responsive enough to detect any sudden changes or potential systemic risk in the market,
especially in the case of an unprecedented event such as the global pandemic. Another paper proposed combin-
ing the two analytical techniques: In that study, documents were semantically analysed and a similarity network
was subsequently constructed47; however, the research focused mainly on the visualisation perspective and was
therefore not correlated with financial markets for the prediction of market volatility. Given the gap in the existing
literature, this paper intends to make two contributions. First, it explores the possibility of quantifying the
topological features of extracted topics dynamically. We adopted a rolling window method to obtain updated
topological results when new articles are provided; this way we could continuously keep track of and understand
how connected the networks were and therefore detect any abnormal patterns immediately. Second, the results
might serve as an indicator, on top of other existing financial indices, that topological features could explain
systemic risk and opportunity in financial markets in the face of a tremendous change or event.

We used the 2015–2016 stock market selloff and COVID-19 as use cases to demonstrate our DTN approach
using all U.S. news articles from Reuters between 1 January 2015 and 31 December 2016, and between 1 January
2019 and 30 September 2020. After data cleaning, we were left with 2,123,284 news articles and 290,771 unique
words. The results revealed significantly less connected topic networks in mid May and November 2015, a few
months before the downfalls of DJIA in August 2015 and January 2016 during the stock market selloff, and in
mid- and late-March 2020, around and after the time when COVID-19 was declared a global pandemic. After
exhibiting an abnormal pattern in March and May 2015, network connectedness showed a gradual rebound to
normal levels. In March 2020, after exhibiting an abnormal pattern for more than a week, network connected-
ness reverted to its original level. This might indicate an increase in investors’ confidence in the financial market,
which was reflected by the gradual pick up of the DJIA a few months or a few days later. The results also showed
the possibility of quantifying and visualising the influence of terms, in particular the emerging term coronavirus
in mid-March 2020 in our use case, on topic similarity networks in different time periods.

Figure 1 shows the design of this research with its six main steps. Step 1 involved collecting news articles
of interest for the analysis. In step 2, the input articles were cleaned, lemmatised, and had repeating phrases
removed; they were then processed using a sliding window method in step 3. Each article subset was semanti-
cally modelled in step 4, and the results were further analysed by constructing topic similarity networks in step
5. Lastly, in step 6, we visualised the final results and compared them to financial indices.

Results
After the standard text cleaning process, lemmatisation, and the removal of repeated phrases, news articles
were sliced by a sliding window of 7 days with a stride of 1 day for dynamic news retrieval to obtain \( M_t \) documents
time \( t \). Each segment of the \( M_t \) documents was semantically analysed using the Latent Dirichlet Allocation
(LDA) model, with the number of topics being set to \( K = 80 \). These topics, each represented by words in the
corpus and their respective topic–word probability matrices, \( \beta_t \), served as the input to form DTNs. We used the
similarity between \( \beta_t \) and \( \beta_j \) to investigate the topological features of the financial news over time, where \( \beta_t \)
is the \( i \)-th topic’s word distribution at time \( t \). The topological features were summarised using network statistics
which can provide insights on financial market evolution, market variation, and systemic risk in financial mar-
kets. More detailed description of the construction of the DTNs can be found in the “Methods” section. The
results are presented and visualised in the following subsections.

Visualisation of dynamic topic networks. To explore the network structures of latent topics from finan-
cial news, we present network diagrams where network nodes represent latent topics and edges are formed by the
similarity of \( \beta_t \) among 80 topics. As an illustration, Fig. 2 shows three network structures at different dates,
where Fig. 2a–c represent the fifteenth day of January, March, and May 2020, respectively. The first date is before
the WHO declared COVID-19 a pandemic on 11 March 2020, and the last two dates are after the announce-
ment. In the figures, each node represents a topic, whose importance is reflected by its node size, proportional
to the number of words generated by the topic across the entire corpus or proportional to the topic probabilities
in $\theta_i$. If the topic contains the word *coronavirus*, its within-topic frequency percentage is obtained from $\beta_{it}$ and represented by the red gradient according to the percentage.

In Fig. 2a, which shows the topic network on 15 January 2020, it is observed that topics were generally highly connected, without any main deviation of nodes from the main cluster. At this time, the pandemic was not an influential subject, therefore we do not see an influence of COVID-19 on any topics and all nodes are of the same colour. Figure 2b for the topic network on 15 March 2020, on the other hand, shows that the network structure was comparatively less connected and some important topics (with big node sizes) were noticeably farther away from the main cluster. This is not surprising as the nodes with different colours suggest that these topics are coronavirus related. With the emergence of this new pandemic-related word and the sudden growth of its influence on news articles, the network structure was altered from dense networks in January 2020 to networks with multiple small clusters in March 2020, causing a sudden drop in network connectedness, as seen in Fig. 2b. Moving forward to 15 May 2020, we observe in Fig. 2c that the term coronavirus still existed in various topics, but it no longer showed a significant influence in the news articles, as it did in March 2020, but instead was integrated with other terms. We can also see that the coronavirus-related nodes have lighter colours compared to those on 15 March 2020, suggesting a much lower influence of the term within topics as well. The network structure as a whole appears similar to that in January 2020 (Fig. 2a).

**Figure 1.** Research design.

**Figure 2.** Network graphs at three different dates. Node size represents topic importance, which is proportional to the number the tokens generated by the topic across the whole corpus. For example, a size of 0.2 represents that an estimated 20% of tokens was covered by the chosen topic $i$. If the topic is coronavirus-related, its node will be represented by a red gradient, whose colour is mapped to the range between 0 and 0.1 according to the word-topic probability for coronavirus in $\beta_{it}$.
Network connectedness. To understand how connected the latent topics were, Figs. 3 and 4 show two network statistics, average degree ($D_t$) and average clustering coefficient ($C_t$), as presented in the “Methods” section, between 1 January 2015 and 31 December 2016, and 1 January 2019 and 30 September 2020. The average degree, $D_t$, in the networks measures how many edges (out of 79, excluding itself) on average link to a node. The higher the average degree, the more dense the network at time $t$ will be. Similarly, the average clustering coefficient, $C_t$, measures how big the clusters formed by neighbourhoods of a node of the network are on average at time $t$. In our context, both $D_t$ and $C_t$ measure the level of connectedness of the topic network at time $t$.

In Fig. 3, the connectedness reached the lowest around late March 2015, with the lowest points 20 March ($D_t = 43.8$, $C_t = 0.783$) and 28 March ($D_t = 32.125$, average $C_t = 0.723$). After around a month of resuming to the normal level, the connectedness began to go downhill again in late April/early May 2015 and reached another local minimum on 12 May ($D_t = 40.25$, $C_t = 0.781$). Moving forward, the connectedness reached yet again another local minimum in mid-November, and the drop was in particular obvious for the $C_t$ ($D_t = 49.175$, $C_t = 0.807$). There were also some moderately low points in late April/early May, early August and early November in 2016.

As observed in Fig. 4, topics were highly connected (with $D_t$ of around 70) at all time $t$ except for March 2020. The two network statistics, $D_t$ and $C_t$, began to show a downward trend on 10 March 2020, and the trend continued until the statistics reached the lowest point on 18 March 2020 for average clustering coefficient, ($C_t = 0.657$) and on 19 March 2020 for average degree ($D_t = 24.475$). An interesting phenomenon is that $D_t$ and

![Figure 3. Network statistics, 1 January 2015 to 31 December 2016.](image)

![Figure 4. Network statistics, 1 January 2019 to 30 September 2020.](image)
C\textsubscript{t} rebounded quickly to “normal levels” after a sharp drop in mid-March 2020. This structural change in topic network connectedness in March 2020 indicates that semantic features summarised by the DTNs constructed by financial news may give us insights into how financial markets react to unexpected events or even catastrophes. Therefore, studying the relationship between these DTNs and financial market movement may provide systemic risk implications from financial news.

Given the results from the network statistics, we try to understand what the topics were at several time points in the pre-pandemic period, the period when the network statistics started to drop to a low value, and the period when the network statistics returned to a normal level. The fifteenth day of January, March, May, and July 2020 were studied, and the top 30 most salient words are shown in Table 1. A majority of the words are finance related, and some of these words, such as *share*, *target*, and *rating*, are in the list of the most salient words in all four randomly selected dates. One interesting finding was that the word *coronavirus*, which was not the most salient word on other dates, topped the ranking on 15 March 2020, a few days after the WHO declared COVID-19 a global pandemic.

As mentioned, the word *coronavirus* was one of the most significant words on 15 March 2020, and thus we show the top four most important COVID-19 related topics on that date in Fig. 5 to understand the underlying word distribution for topic \textit{i}, i.e. $f_{it}$. The four topics are topics 1, 3, 4, and 6, where the first three topics contain the word *coronavirus* as a major word and topic 6 contains *travel* and *flight* as major words. Topics are represented by bubbles, whose sizes are proportional to their importance and which are clustered together on the basis of their similarities. The four topics, 1, 3, 4, and 6 are highlighted by red bubbles in four diagrams. We can see that *coronavirus* had a significant importance for the selected news period (9–15 March 2020). Also, it comes as no surprise that the bubbles corresponding to the four most important topics are close to each other, indicating a relatively higher across-topic similarity. Topic 1 is related to the financial market in general and how it reacted to the global pandemic. Topic 3 focuses on the societal perspective and the latest development of the virus. Topic 4, on the other hand, links the pandemic to politics in the United States, and Topic 6 shows how coronavirus was affecting the world and the travel industry. The above topic characteristics explain why the connectedness

| Rank | 15 Jan 2020 | 15 Mar 2020 | 15 May 2020 | 15 Jul 2020 |
|------|-------------|-------------|-------------|-------------|
| 1    | Share       | Target      | Rating      | Target      |
| 2    | Rating      | Share       | Price       | Rating      |
| 3    | Target      | Market      | Share       | Price       |
| 4    | Side        | Price       | Side        | Share       |
| 5    | Order       | Cut         | Target      | Credit      |
| 6    | Price       | March       | Inc         | Inc         |
| 7    | Inc         | Nasdaq      | Order       | Nyse        |
| 8    | China       | Bank        | Company     | Corp        |
| 9    | Corp        | Coronavirus  | Corp        | Order       |
| 10   | Year        | Rating      | Credit      | Fitch       |
| 11   | Bank        | Company     | Fund        | Trump       |
| 12   | Group       | Report      | Security    | Side        |
| 13   | Trade       | Inc         | Bank        | Imbalance   |
| 14   | Stock       | World       | Source      | Company     |
| 15   | Trump       | Service     | China       | Security    |
| 16   | Report      | Security    | Trump       | Stock       |
| 17   | Company     | Trump       | Service     | Group       |
| 18   | Trial       | Oil         | Group       | China       |
| 19   | Deal        | Event       | Year        | State       |
| 20   | Revenue     | Credit      | State       | Bank        |
| 21   | Agreement   | Season      | Report      | Source      |
| 22   | Tuesday     | Rate        | Stock       | Holding     |
| 23   | Credit      | Business    | Thursday    | Year        |
| 24   | Cut         | Thursday    | Week        | Trade       |
| 25   | Service     | Corp        | Market      | June        |
| 26   | State       | Week        | Cash        | Market      |
| 27   | Energy      | Energy      | Case        | President   |
| 28   | Market      | Year        | Rate        | Investment  |
| 29   | Security    | April       | Revenue     | July        |
| 30   | Sale        | China       | Usd         | Case        |

Table 1. Top 30 most salient words for the fifteenth day of January, March, May, and July 2020. The term *coronavirus* is the most salient term only on 15 March 2020.
of the topic network can be particularly low when COVID-19, as an emerging global event, tended to have a big influence on financial news.

**Relationships with financial indices.** We chose to relate our dynamic topic network statistics to two indices: (1) DJIA: Composed of the 30 largest stocks listed on the U.S. stock exchanges, the DJIA is a good representative of U.S. stock market performance; (2) VIX: This is a derived index from S&P 500; it is widely used by investors and financial institutions to measure market risk and stability and is a good proxy for systemic risk in financial markets.

A comparison of the network statistics with the VIX and DJIA gave us some interesting results. Figures 6 and 7 display the network statistics and the DJIA in the same graph. It can be seen that the topic network statistics, $D_t$ and $C_t$, showed two main drops around April/May and November 2015, several months before the drops of DJIA in September 2015 and January 2016, respectively and a general downward trend that aligned with the DJIA in March 2020. During the COVID-19 pandemic, the two network statistics took a sudden dip for a few days and then quickly rebounded to their original levels and became stationary again, whereas the DJIA reached the lowest point of 18,591.93 on 23 March 2020 before gradually recovering to its original level.

Given the patterns in Figs. 6 and 7, one might be curious to see if there is any correlation or lead-lag relationship between the time series of the network statistics and the DJIA. Since we were also interested in whether one series was ahead of another, we needed a way to quantify the synchrony between the two series. We adopted the dynamic time warping (DTW) algorithm to detect the leader-follower interaction. Specifically, we used the DTW algorithm to evaluate the similarity between network statistics and DJIA. It starts with the first time step of the query series and calculates its distance with all data points on the reference series. Then it moves to the second time step of the query series and it repeats the same calculation until the last time step. The best alignment between the two series is found by getting the minimum warping path using the calculations. Figures 8 and 9 show the point-by-point comparison between the average degree, $D_t$, and the DJIA for the years 2015–2016 and 2020 respectively. The grey lines connect the two series on the basis of the shortest Euclidean distance with the

---

**Figure 5.** Top 4 most important **coronavirus**-related topics on 15 March 2020.
Figure 6. Comparison of network statistics with DJIA, 1 January 2015 to 31 December 2016.

Figure 7. Comparison of network statistics with DJIA, 1 January 2019 to 30 September 2020.

Figure 8. Dynamic time warping: point-by-point comparison between average degree and DJIA for the years 2015 and 2016, both series normalised.
implementation of an asymmetric step pattern\textsuperscript{51}. In terms of the major events, Fig. 8b,c show that the downturns of average degree in May and November 2015 correspond to the low points of DJIA in early September 2015 and January 2016. Some moderate fluctuations also occurred in 2016, but the similarity relationship was not very obvious. In Fig. 9b, when the average degree dropped to the lowest point on 18 March 2020 and began the pick up to the original level, it provided an indication that the DJIA would also gradually start to pick up in the upcoming months. The direction of the grey lines before 18 March 2020 seems to tell us that the average degree reacted faster than the DJIA. After 18 March 2020, the grey lines showing the same direction as just before 18 March 2020 are even steeper and show stronger evidence that the average degree “led” the DJIA to return to normal levels. The responsive nature of the network statistics potentially provides insights for academics, investors, and financial institutions on when the market would start to recover after an unprecedented event, such as the COVID-19 pandemic investigated in this study.

Figure 10 shows another comparison between the average degree of the topic networks, VIX, and DJIA in 2020. These three series experienced an unusual movement during the pandemic period in February and March 2020. VIX showed an increasing trend starting in late February 2020, while the DJIA exhibited a substantial drop at a similar time until it reached its lowest point of 18591.93 on 23 March 2020. The two topic network statistics, $D_t$ and $C_t$, started to show a drastic decrease in early March, dropping to their lowest value on 18 March 2020 before responsively returning to their original value a few days later.

**Discussion**

The severe downturn of financial markets or systemic risk can be attributed to different events or financial incidents. A possible incident is related to credit markets and a substantial increase in default probabilities in a short period, for example, in the financial tsunami in 2008. There is common scientific evidence of having higher systemic risk from a significant increase in financial network connectedness\textsuperscript{45,52–54}. The idea of using the topic network connectedness of latent topics to capture possible systemic risk is explained in Fig. 2. When an event that is of great financial relevance is evolving, there can have a transition in the topic network structure from
methods to understand news articles but from other perspectives. This paper looks at the methodology from markets, especially on systemic risk management. It is important to note that the DTW algorithm tries to match detecting opportunities in the financial market. The dynamic information in news articles can supplement other DJIA would also slowly pick up in the coming month. This further highlights the possibility of using news 19 pandemic, When the average degree began to recover in mid-March 2020, it provided an indication that the DJIA (i.e. possibly showing a leading time series signal for future movement of the DJIA). During the COVID-19 pandemic, Possible channels of the transmission can also be found in. The implication of our methodology does not present evidence that the pandemic network connectedness can lead the financial network connectedness. The above research documented the possible contagion or transmission of the pandemic risk to financial risk. To further elaborate the use of the dynamic topic networks as a possible tool to provide early warning signals for severe financial downturns, we have included the use cases of the 2015–2016 stock market selloff and the COVID-19 pandemic. We find a substantial drop in the topic network connectedness in March and May 2015 before seeing a substantial adjustment in the DJIA in August 2015. We also see a sudden drop in the topic network connectedness in March 2020 during the COVID-19 pandemic.

To understand the underlying topics in news articles and whether they could provide an early assessment of systemic risk in financial markets, we made use of LDA for topic modelling, and the latent topics were used to construct DTNs on the basis of topic similarities. The final outputs (i.e. metrics of network connectedness) were related to benchmark market indices to investigate the predictive power of our approach. Using the 2015–2016 stock market selloff as a case, Fig. 3 reveals a sudden drop in the topic network connectedness in March and May 2015. These significant decline in topic network connectedness can be an indication of the propagation of fear in financial markets which is reflected in the evolution of new influential topics in the corpus. As seen in Fig. 6, following the significant drops in the topic network connectedness in about three months, we see a steep downturn in DJIA in August 2015. Using COVID-19 as a case having worldwide impacts, we conclude that DTNs and their network statistics could be auxiliary measures to detect systemic risk and, in addition, could provide an opportunity to determine when a market might pick up after a tremendous change in the financial markets due to extraordinary events or catastrophes. As seen in Fig. 4, the topic networks began to be less connected on 10 March 2020. As we used a sliding window length of 7 days, this means the analysis on that date included news articles from 4 to 10 March 2020. This might be an indication that investors already foresaw instability in the financial market on 11 March 2020, and this loss in confidence was captured in the news articles prior to the slump in the DJIA. Slightly more than a week after the global pandemic declaration, the topic network statistics started to pick up again and rapidly returned to their original level. The DJIA exhibited a similar pattern a few days after the recovery of the topic network statistics, the only difference being that the increasing trend was much slower.

As stated in the “Introduction”, VIX is one of the widely recognised indices for measuring market volatility and a main risk indicator for numerous financial institutions and investors to refer to, although some have argued that it might have limited predictive power. In this research, we included VIX for comparison with the network analysis results. As can be seen in Fig. 10, VIX started to drop at a similar time to when the network statistics rose. Same as the DJIA, however, the rate was comparatively slow due to the fact that it was highly related to the derivative products on the S&P500 index, and it was not until May 2020 that VIX began to reach a plateau. The topic network statistics are much more responsive measures compared to VIX, possibly because news articles are real-time information that contain a variety of insights and latest events from around the globe. Financial news, in particular, is a useful source that captures and encapsulates investors’ confidence in the market. The increase in the topic network statistics a week after the global pandemic declaration might have provided an indication that investors were experiencing less uncertainty and becoming less anxious about future movement in the financial market, which could therefore subsequently explain the gradual increase in the DJIA that followed a few days later.

The point-by-point comparison in Figs. 8 and 9 also suggest that the average degree, \( D_t \), was ahead of the DJIA (i.e. possibly showing a leading time series signal for future movement of the DJIA). During the COVID-19 pandemic, When the average degree began to recover in mid-March 2020, it provided an indication that the DJIA would also slowly pick up in the coming month. This further highlights the possibility of using news articles to ascertain when the recovery of global stock markets might happen after a period of crash and thus detect opportunities in the financial market. The dynamic information in news articles can supplement other indices, such as VIX, in guiding investors and financial institutions to have better predictive power in financial markets, especially on systemic risk management. It is important to note that the DTW algorithm tries to match the patterns by constructing the optimal warping paths and does not necessarily imply causality. Nevertheless, it provides us with a glimpse of the similarity between the two series and how time indices in one series could be matched to those in the other series.

This paper shows the possibility of combining the topic modelling and network analysis of textual information to give predictions in the financial market. A limited number of studies have attempted to combine the two methods to understand news articles but from other perspectives. This paper looks at the methodology from the perspective of constructing DTNs and uses topic network statistics to capture network connectedness hidden in text and financial news. We also relate our topic network results to the DJIA and VIX. The results show the potential of using this novel approach to predict the influence of the pandemic on financial risk. The COVID-19 pandemic induced pandemic-related salient words in the news articles on 15 March 2020, and it is evident from Fig. 2 that their presence caused topic similarity networks to be less coherent. The word coronavirus was still in the corpus after two months; however, it was much less influential and began to become part of other bigger topics. Moreover, our proposed topic network approach allowed us to easily quantify the influence of latent topics dynamically, as in Fig. 4. Through the combination of topic modelling and network analysis, we have gained a better understanding of the influential words and topics in selected time periods.

This paper takes the COVID-19 pandemic as a use case. The paper documented the extraordinarily high financial network connectedness due to the impacts of the COVID-19 pandemic. In addition, the paper presented evidence that the pandemic network connectedness can lead the financial network connectedness. The above research documented the possible contagion or transmission of the pandemic risk to financial risk. Possible channels of the transmission can also be found in. The implication of our methodology does not limit to the COVID-19 pandemic. To illustrate further insights of the dynamic topic networks on financial risk
Table 2. Examples of news articles.

| Posting date | Headline                                                                 | Body text                                                                 |
|--------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------|
| 1 Jan 2020   | A major shake-up for oil and shipping                                      | Tougher rules on sulphur emissions from ships came into effect on Wednesday, in the biggest shake-up for the oil and shipping industries for decades. |
| 1 Jan 2020   | U.S. auto safety agency to probe fatal Tesla crash in Los Angeles         | The U.S. National Highway Traffic Safety Administration (NHTSA) said late on Tuesday it will investigate a fatal Dec. 29 Tesla Inc. (> TSLA.O ) crash in Los Angeles that killed two people |
| 1 Jan 2020   | Brookfield Ontario 189 MW Prince wind power plant returns to service—Ontario IESO | Brookfield Ontario 189 MW Prince wind power plant returns to service, the Ontario IESO said on Tuesday |
| 1 Jan 2020   | Iran's Khamenei strongly condemns U.S. attacks in Iraq—TV                 | Iran's Supreme Leader Ayatollah Ali Khamenei strongly condemned U.S. attacks on Iran-allied militia group in Iraq, Iranian state TV reported on Wednesday. |

Methods

This section describes the methodology of the research. Each step corresponds to the research design in Fig. 1. Our data analysis is entirely based on Python. We used gensim for most of the natural language processing, in particular the LDA that we heavily focused on in this research. To create network graphs, we used networkx. To create topic distribution graphs in Fig. 5, we used pyLDAvis. For the DTW algorithm, we used dtw-python. All other visualizations are made with the package matplotlib.

Data. For this research, we collected U.S. news articles between 1 January 2015 and 31 December 2016, and 1 January 2019 and 30 September 2020 from Reuters. Table 2 shows a few examples of financial news with the posting date, headline, and body text.

Text cleaning. For each article, the headline and the body text were combined to form the full text. Standard text-cleaning procedures were carried out, including the removal of white spaces, non-alphabetical characters, and stop-words and the conversion of all words to lower case. Only nouns and proper nouns (part-of-speech tags: NOUN and PROPN)56 were retained before applying lemmatisation. Words that appeared only once in a given period are removed from the constructed dictionary. On the other hand, there were numerous repeated phrases in the articles: for instance, 'Reporting by ... ', 'For company coverage: ... ' , and 'Click the following link to watch video: ... ' , and so on. These phrases do not add value to the analysis and were therefore removed using regular expressions. The final cleaned version consisted of 2123284 news articles with 290771 unique words, and this served as the input for the topic modelling. Tables 3, 4, 5, and 6 show the number of articles and unique words per month for 2015, 2016, 2019, and 2020, respectively.

Sliding window. This study adopted a sliding window approach in which news articles in a given window length were extracted, semantically modelled, and analysed with a topic similarity network; then, the window would slide to the next period. These steps were repeated until the window reached the last available date of the provided news articles. In this study, the window length was set as 7 days with a stride of 1 day so that the LDA in the next step could capture the latent topics with enough news articles while also being able to reflect changes in a responsive manner. Specifically, on day t, we used the documents on day t − 6 to day t to perform the LDA. By doing so, we intended to observe how the connectedness between latent topics changed over time.
Latent Dirichlet Allocation. Topic modelling was used to understand the latent topics behind news articles. Among the various methods available, LDA is widely used\(^5\). This generative probabilistic model represents a document as a random mixture of latent topics, where each topic is further characterised by a distribution of words. Suppose we have \(N\) words in the corpus and \(M_t\) documents on days \(t - 6\) to \(t\). Given \(M_t\) documents as input at time \(t\), let \(K\) be the number of latent topics in the documents. The \(M_t\) documents contain a sequence of words which are assumed to be generated from a Dirichlet mixture of multinomial distributions under LDA in two steps. For each document at time \(t\), the first step is to generate a latent topic probability vector \(\theta_t\) from a \(K\)-dimensional Dirichlet distribution, denoted by \(\text{Dirichlet}_K(\alpha_t)\) with a \(K \times 1\) parameter vector \(\alpha_t\). In the second step, given the latent topic probability vector \(\theta_t\), we first simulate a latent topic vector \(z_n\) for the \(n\)-th word. Then, for \(z_n = i\), where \(i\) is a topic index between 1 and \(K\), we generate the \(n\)-th word from the multinomial distribution with the \(N \times 1\) probability vector \(\beta_{it}\). Combining all \(K\) vectors \(\beta_{it}, i = 1, \ldots, K\), we can form a \(K \times N\) topic-word probability matrix at time \(t\). We are interested in estimating \(\alpha_t\) and \(\beta_t\) using LDA to determine the topological features of the latent topics of financial news.

---

**Table 3.** Number of articles and unique words in 2015.

| Month   | Number of articles | Number of unique words | Month   | Number of articles | Number of unique words |
|---------|--------------------|------------------------|---------|--------------------|------------------------|
| 2015/01 | 63,193             | 49,834                 | 2015/07 | 68,813             | 53,002                 |
| 2015/02 | 66,933             | 51,453                 | 2015/08 | 60,503             | 49,297                 |
| 2015/03 | 73,658             | 53,916                 | 2015/09 | 62,556             | 49,616                 |
| 2015/04 | 70,974             | 53,780                 | 2015/10 | 71,587             | 52,612                 |
| 2015/05 | 67,986             | 53,293                 | 2015/11 | 67,638             | 52,347                 |
| 2015/06 | 68,767             | 52,812                 | 2015/12 | 51,148             | 46,441                 |

**Table 4.** Number of articles and unique words in 2016.

| Month   | Number of articles | Number of unique words | Month   | Number of articles | Number of unique words |
|---------|--------------------|------------------------|---------|--------------------|------------------------|
| 2016/01 | 57,453             | 40,006                 | 2016/07 | 61,567             | 38,709                 |
| 2016/02 | 65,954             | 41,828                 | 2016/08 | 61,607             | 40,114                 |
| 2016/03 | 62,898             | 41,377                 | 2016/09 | 53,616             | 38,768                 |
| 2016/04 | 65,097             | 40,047                 | 2016/10 | 55,694             | 37,642                 |
| 2016/05 | 66,997             | 40,685                 | 2016/11 | 60,381             | 39,336                 |
| 2016/06 | 64,077             | 39,685                 | 2016/12 | 46,552             | 34,559                 |

**Table 5.** Number of articles and unique words in 2019.

| Month   | Number of articles | Number of unique words | Month   | Number of articles | Number of unique words |
|---------|--------------------|------------------------|---------|--------------------|------------------------|
| 2019/01 | 27,932             | 24,543                 | 2019/07 | 27,876             | 25,024                 |
| 2019/02 | 26,776             | 24,419                 | 2019/08 | 29,152             | 24,424                 |
| 2019/03 | 29,819             | 25,294                 | 2019/09 | 27,810             | 24,451                 |
| 2019/04 | 27,495             | 24,368                 | 2019/10 | 31,244             | 26,356                 |
| 2019/05 | 29,827             | 25,848                 | 2019/11 | 28,577             | 24,420                 |
| 2019/06 | 27,464             | 23,795                 | 2019/12 | 24,305             | 21,665                 |

**Table 6.** Number of articles and unique words in 2020.

| Month   | Number of articles | Number of unique words | Month   | Number of articles | Number of unique words |
|---------|--------------------|------------------------|---------|--------------------|------------------------|
| 2020/01 | 27,942             | 24,096                 | 2020/07 | 27,965             | 23,412                 |
| 2020/02 | 27,307             | 23,607                 | 2020/08 | 26,583             | 22,434                 |
| 2020/03 | 43,727             | 27,872                 | 2020/09 | 38,536             | 23,851                 |
| 2020/04 | 31,367             | 23,484                 |         |                    |                        |
| 2020/05 | 23,707             | 21,672                 |         |                    |                        |
| 2020/06 | 32,224             | 23,359                 |         |                    |                        |
In an unsupervised learning setting, there is no ground truth of how many topics \( K \) are behind the articles. Literature has suggested the use of a large number. In previous studies, \( K \) was arbitrarily set to a large number\(^5\). In this paper, we selected 80 as the number of topics.

**Topic similarity network.** The primary focus of network analysis is to better understand relationships between entities. This study intended to take the output from the LDA model as the input into the topic similarity networks so that we could further understand topic connectedness and how it changes over time.

**Measuring topic similarity.** As previously mentioned, topics are represented as distributions of words and captured in the matrix \( \beta \). Therefore, the similarity between two topics can be measured by comparing their word distributions. Kullback-Leibler (KL) divergence is often used when calculating the similarity between distributions. Although it is intuitively used to measure the distance between probability distributions, it is not a true metric given its asymmetry and the fact that it does not satisfy the triangle inequality. Alternatively, this study adopted Jensen-Shannon (JS) divergence, which is based on KL divergence, to calculate a normalised score. In other words, JS divergence can be considered as a symmetric and smoothed version of KL divergence. Equation (1) shows how JS divergence between probability distributions \( P \) and \( Q \) is calculated:

\[
JS(P||Q) = \sum \frac{1}{2} P \times \log \frac{P}{M(P,Q)} + \frac{1}{2} Q \times \log \frac{Q}{M(P,Q)},
\]

where

\[
M(P,Q) = \frac{1}{2} (P + Q).
\]

When \( P = Q \), the JS divergence will be zero. The smaller the \( JS(P||Q) \), the more similar \( P \) and \( Q \) will be.

**Topic network.** A topic similarity network \( G_t = (V_t, E_t) \) can be constructed, where \( V_t \) is the set of \( K \) vertices representing \( K \) topics at time \( t \) and \( E_t \) is the set of edges at time \( t \). As previously mentioned, \( K \) equals 80 in this paper. As shown in Eq. (3), an edge between two topics \( e_{ijt} \) for \( 1 \leq i,j \leq n \) and \( i \neq j \) at time \( t \) is established if their divergence \( JS(\beta_{it}||\beta_{jt}) \) in terms of topic-word distribution is below a certain predefined threshold, \( \kappa \). In this paper, we have included a general guideline which is similar to an ‘empirical rule’ in Statistics to filter out abnormally large JS divergence values. After constructing the dynamic topic networks, we propose using the mean and standard deviation of the median JS divergence of distance pairs to define the threshold, \( \kappa \). Our data analysis on the two cases of the 2015–2016 stock market selloff and the COVID-19 pandemic indicates that setting \( \kappa \) at three standard deviations above the mean highlights topological properties of the dynamic topic networks which are able to generate early warning signals of severe financial market downturns or systemic risk. As in an early paper\(^4\), we have also conducted the sensitivity analysis to study the effect of the threshold on the topic network connectedness. Figure 11 shows the result during the 2015–2016 stock market selloff. The signals from using different thresholds are largely consistent in the sense that we see obvious drops in the average degree of the topic networks in March and May 2015, though the empirical rule of ‘three standard deviations above the mean’ may give a better result or a clear early warning signal of severe financial adjustments. The sensitivity analysis gives us extra confidence in using the dynamic topic network methodology as an additional tool for assessing systemic risk.

The data analysis suggested that the median JS divergence of distance pairs given a week of news articles had an mean of 0.73 and a standard deviation of 0.009 for the 2015–2016 time series, and those of 0.807 and 0.0056 for the 2019–2020 series. We performed a sensitivity analysis to understand how the average degree of networks would change and suggested the empirical rule. In this paper, we present results by setting \( \kappa \) to three standard deviations above the mean, i.e. 0.76 for the 2015–2016 stock market selloff and 0.82 for the COVID-19 pandemic.

![Figure 11. Threshold sensitivity for the period of 2015 to 2016.](https://doi.org/10.1038/s41598-022-06399-x)
\[ e_{ijt} = \begin{cases} 1 & \text{if } JS(\beta_i || \beta_j) \leq \kappa, \\ 0 & \text{otherwise}. \end{cases} \]

Average degree \( D_t \) as shown in Eq. (4) is a straightforward measure to calculate how many edges \(|E_t|\) are in \( G_t \) with \(|V_t|\) vertices.

\[ D_t = \frac{2|E_t|}{|V_t|} = \frac{2|E_t|}{K}. \]

To measure the degree to which the nodes tend to be linked together at time \( t \), the average clustering coefficient \( C_t \) is calculated. For any vertex \( i \), its neighbourhood \( H_{it} \) can be defined as \( H_{it} = \{ j : e_{ijt} \in E_t \} \). The number of vertices of \( H_{it} \) is defined as \( N_{it} \). Therefore, the local clustering coefficient \( C_{it}^{\text{local}} \) for vertex \( i \) is calculated as

\[ C_{it}^{\text{local}} = \frac{2|\{ e_{jkt} : j, k \in H_{it}, e_{jkt} \in E_t \}|}{k_{it}(k_{it} - 1)}. \]

The average clustering coefficient \( C_t \) at any given time is simply the average of the local clustering coefficients of all vertices in \( G_t \):

\[ C_t = \frac{1}{K} \sum_{i=1}^{K} C_{it}^{\text{local}}. \]

**Visualisation.** pyLDAvis was used to help visualise the topic model in an interactive manner. Topics were projected on a two-dimensional space using principal coordinates analysis, in which the distance matrix was created by calculating the JS divergence between topic-term distributions \( \beta_t \). The importance of topic \( k \) is represented by its node size, \( A_k \propto N_k / \sum_k N_k \), proportional to the number of tokens generated by the topic across the corpus, with \( N_k = \sum_d n_{dk} N_d \), where \( N_d \) denotes the number of words in document \( d \) and \( d = 1, 2, ..., M_t \). There is a relevance metric \( \lambda \) to control the ranking of terms. It has a range of \( 0 \leq \lambda \leq 1 \), with 1 indicating terms are ranked on the basis of their topic-specific probability in a descending order and 0 meaning terms are entirely ranked by their \( \beta_t \), a term’s topic-specific probability over its marginal probability across the whole corpus. In this paper, \( \lambda \) is set to 0.6 in accordance with the literature.

**Data availability**

The datasets generated during and/or analysed during the current study are not publicly available due to the license of the datasets used, but are available from the corresponding author on reasonable request.

Received: 9 July 2021; Accepted: 28 January 2022

Published online: 17 February 2022

**References**

1. Eisenberg, L. & Noe, T. H. Systemic risk in financial systems. *Manag. Sci.* 47, 236–249 (2001).
2. Kaufman, G. G. & Scott, K. E. What is systemic risk, and do bank regulators retard or contribute to it?. *Indep. Rev.* 7, 371–391 (2003).
3. Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* 469, 351–355 (2011).
4. Wikipedia contributors. 2015–2016 stock market selloff—Wikipedia, the free encyclopedia (2021) (Accessed 31 Oct 2021).
5. Zhong, X. & Raghib, M. Revisiting the use of web search data for stock market movements.
6. Wikipedia contributors. 2015–2016 stock market selloff—Wikipedia, the free encyclopedia (2021) (Accessed 31 Oct 2021).
7. Baumöhl, E., Bouri, E., Hoang, T.-H.-V., Shahzad, S. J. H. & Výrost, T. From physical to financial contagion: the COVID-19 pandemic and increasing systemic risk among banks. *Tech. Rep., Kiel, Hamburg*. (2020).
8. Whaley, R. E. Understanding the VIX. *J. Portfolio Manag.* 35, 98–105. https://doi.org/10.3905/JPM.2009.35.3.098 (2009).
9. Wang, J., Lu, X., He, F. & Ma, F. Which popular predictor is more useful to forecast international stock markets during the coronavirus pandemic: VIX vs EPU? *J. Int. Finan. Anal.* 72, 101596 https://doi.org/10.1016/j.jinaf.2020.101596 (2020).
10. Bekker, G. & Hoerova, M. The VIX, the variance premium and stock market volatility. *J. Econom.* 183, 181–192. https://doi.org/10.1016/j.joec.2014.05.008 (2014).
11. Bekker, G., Hoerova, M. & Lo Duca, M. Risk, uncertainty and monetary policy. *J. Monet. Econ.* 60, 771–788. https://doi.org/10.1016/j.jmoneco.2013.06.003 (2013).
12. Chow, K. V., Jiang, W. & Liu, J. Does VIX Truly Measure Return Volatility?, chap. 40, 1533–1559 (Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning. (2018). https://doi.org/10.1142/9789811202391_0040).
13. Battiston, S., Puliga, M., Kaushik, R., Tasca, P. & Caldarelli, G. Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Sci. Rep.* 2, 1–6 (2012).
14. Thurner, S. & Poledna, S. Debtrank: transparency: Controlling systemic risk in financial networks. *Sci. Rep.* 3, 1–7 (2013).
15. Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* 469, 351–355 (2011).
16. Chu, C., Vodenska, I., Havlin, S. & Stanley, H. E. Cascading failures in b-inpartite graphs: Model for systemic risk propagation. *Sci. Rep.* 3, 1–9 (2013).
17. Cimini, G., Squartini, T., Garlaschelli, D. & Gabrielli, A. Systemic risk analysis on reconstructed economic and financial networks. *Sci. Rep.* 5, 1–12 (2015).
18. Delphini, D., Battiston, S., Caldarelli, G. & Riccaboni, M. Systemic risk from investment similarities. *PLoS ONE* 14, e0217141 (2019).
19. Song, J., Zhang, Z. & So, M. K. P. On the predictive power of network statistics for financial risk indicators. *J. Int. Finan. Mark. Inst. Money* 75, 101420 (2021).
20. Meng, H. et al. Systemic risk and spatiotemporal dynamics of the US housing market. Sci. Rep. 4, 1–7 (2014).

21. Kou, G., Chao, X., Peng, Y., Alsaadi, F. E. & Herrera-Viedma, E. Machine learning methods for systemic risk analysis in financial sectors. Technol. Econ. Dev. Econ. 25, 716–742 (2019).

22. Zheng, Z., Podobnik, B., Feng, L. & Li, B. Changes in cross-correlations as an indicator for systemic risk. Sci. Rep. 2, 1–8 (2012).

23. WHO director-general’s opening remarks at the media briefing on COVID-19—11 March 2020.

24. Zhang, D., Hu, M. & Ji, Q. Financial markets under the global pandemic of COVID-19. Finance Res. Lett. 36, 101528 https://doi.org/10.1016/j.frl.2020.101528 (2020).

25. Albulescu, C. T. COVID-19 and the united states financial markets’ volatility. Finance Res. Lett. 38, 101699 https://doi.org/10.1016/j.frl.2020.101699 (2021).

26. Sansa, N. A. The impact of the COVID-19 on the financial markets: Evidence from China and USA. Electron. Res. J. Soc. Sci. Humant. https://doi.org/10.2193/srrn.3567901 (2020).

27. Ali, M., Alam, N. & Rizvi, S. A. R. Coronavirus (COVID-19)—An epidemic or pandemic for financial markets. J. Behav. Exp. Finance 27, 100341 https://doi.org/10.1016/j.jbef.2020.100341 (2020).

28. Hisano, R., Sornette, D., Mizuno, T., Ohnishi, T. & Watanabe, T. High quality topic extraction from business news explains abnormal financial market volatility. PLoS ONE 8, 1–12. https://doi.org/10.1371/journal.pone.0064846 (2013).

29. Glasserman, P. & Mamaysky, H. Does unusual news forecast market stress? J. Financial Quant. Anal. 54, 1937–1974. https://doi.org/10.1017/S0022109099000127 (2019).

30. Atkins, A., Niranjan, M. & G erding, E. Financial news predicts stock market volatility better than close price. J. Finance Data Sci. 4, 120–137. https://doi.org/10.1016/j.jfds.2018.02.002 (2018).

31. Morimoto, T. & Yasuhara, Y. Financial forecasting market volatility using a dynamic topic model. Asia-Pac Financial Mark. 24, 149–167. https://doi.org/10.1007/s10690-017-9228-z (2017).

32. Curme, C., Preis, T., Stanley, H. E. & Moat, H. S. Quantifying the semantics of search behavior before stock market moves. Proc. Natl. Acad. Sci. 111, 11600–11605. https://doi.org/10.1073/pnas.1324054111 (2014).

33. Bollen, J., Mao, H. & Zeng, X. Twitter mood predicts the stock market. J. Comput. Sci. 2, 1–8. https://doi.org/10.1016/j.jocs.2010.12.007 (2011).

34. Chen, W., Cai, Y., Lai, K. K. & Xie, H. A topic-based sentiment analysis model to predict stock market price movement using Weibo mood. Web Int. Agent Syst. 14, 287–300. https://doi.org/10.3233/WIA-160345 (2016).

35. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using networkx. In Proceedings of the 7th Python in Science Conference, (eds Varoquaux, G. et al.) 11–15 (2008).

36. Borgatti, S. P., Mehra, A., Brass, D. J. & Labianca, G. Network analysis in the social sciences. Science 323, 892–895. https://doi.org/10.1126/science.1156821 (2009).

37. Bansal, S., Read, J., Pourahmadian, B. & Meyers, L. A. The dynamic nature of contact networks in infectious disease epidemiology. J. Theor. Biol. 4, 478–489. https://doi.org/10.1080/17513758.2010.530376 (2010).

38. Li, J. et al. Application of weighted gene co-expression network analysis for data from paired design. Sci. Rep. 8, 622. https://doi.org/10.1038/s41598-017-18705-z (2018).

39. Kim, H.-H., Zhang, L.-L., Yan, R., Lu, J.-J. & Hu, Y. Network analysis of drug-target interactions: A study on FDA-approved new molecular entities between 2000 to 2015. Sci. Rep. 7, 12230. https://doi.org/10.1038/s41598-017-12061-8 (2017).

40. Elliott, M., Golub, B. & Jackson, M. O. Financial networks and contagion. Am. Econ. Rev. 104, 3115–53. https://doi.org/10.1257/aer.104.10.3115 (2014).

41. So, M. K. P., Tiwari, A., Chu, A. M. Y., Tsang, J. T. Y. & Chan, J. N. L. Visualizing COVID-19 pandemic risk through network connectedness. Int. J. Infect. Dis. 96, 558–561. https://doi.org/10.1016/j.ijid.2020.05.011 (2020).

42. Chu, A. M. Y., Tiwari, A. & So, M. K. P. Detecting early signals of COVID-19 global pandemic from network density. J. Travel Med. 27, taaa084 (2020).

43. Chu, A. M. Y., Chan, J. N. L., Tsang, J. T. Y., Tiwari, A. & So, M. K. P. Analyzing cross-country pandemic connectedness during COVID-19 using a spatial-temporal database: Network analysis. JMI R Public Health Surveill. 7, e27317 (2021).

44. So, M. K. P., Chu, A. M. Y., Tsang, J. T. Y., Tiwari, A. & Chan, J. N. L. On topological properties of COVID-19: Predicting and assessing pandemic risk with network statistics. Sci. Rep. 11, 5112. https://doi.org/10.1038/s41598-021-84094-z (2021).

45. So, M. K. P., Chu, A. M. Y. & Chan, T. W. C. Impacts of the COVID-19 pandemic on financial market connectedness. Finance Res. Lett. 38, 101864 https://doi.org/10.1016/j.frl.2020.101864 (2021).

46. So, M. K. P., Chan, L. S. H. & Chu, A. M. Y. Financial network connectedness and systemic risk during the COVID-19 pandemic. Asia-Pac. Financial Mark. 28, 1–17 (2021).

47. Chu, A. M. Y., Chan, L. S. H. & So, M. K. P. Stochastic actor-oriented modelling of the impact of COVID-19 on financial network evolution. Stat 10, e408 (2021).

48. Kawata, S. & Fujiwara, Y. Constructing of network from topics and their temporal change in the Nikkei newspaper articles. Evol. Inst. Econ. Rev. 13, 423–436. https://doi.org/10.1007/s10484-016-0061-2 (2016).

49. Maiya, A. S. & Rolfe, R. M. Topic similarity networks: Visual analytics for large document sets. In 2014 IEEE International Conference on Big Data (Big Data), 364–372, https://doi.org/10.1109/BigData.2014.7004253 (2014).

50. Savova, K. G., Lesh, N., Crerar, C. & Shiflet, K. LEDavii: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 58–72 https://doi.org/10.3115/v1/W14-3110 (Association for Computational Linguistics, 2014).

51. Toni, G. Computing and visualizing dynamic time warping alignments in r: The dtw package. J. Stat. Softw. 31, 1–24 https://doi.org/10.18637/jss.v031.i07 (2009).

52. Billio, M., Getmanovsky, M., Lo, A. W. & Pelizzon, L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J. Financial Econ. 104, 535–559 (2012).

53. Barunik, J. & Křehlík, T. Measuring the frequency dynamics of financial connectedness and systemic risk. J. Financial Econ. 16, 271–296 (2018).

54. Zhang, D. & Broadstock, D. C. Global financial crisis and rising connectedness in the international commodity markets. Int. Rev. Financial Anal. 68, 101239 (2020).

55. Bie, D. M., Ng, A. Y., Jordan, M. I. & Lafferty, J. Latent Dirichlet allocation. J. Mach. Learn. Res. 3, 2003 (2003).

56. Universal pos tags. http://universaldependencies.org/ pos/. (Accessed 01 Feb 2010).

Acknowledgements

This study was partially supported by the Hong Kong RGC Theme-based Research Scheme, grant number T31-604/18-N and the Hong Kong University of Science and Technology research grant “Big Data Analytics on Social Research” (grant# CEF20BM04).
Author contributions
M.K.P.S. and A.M.Y.C. conceptualized the study. M.K.P.S. and A.S.W.M. collected and analyzed the data with statistical modeling. M.K.P.S., A.M.Y.C. and A.S.W.M. interpreted the results. M.K.P.S. and A.S.W.M. drafted the manuscript. A.M.Y.C. finalized the manuscript. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.K.P.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022