Research Article

Analysis of Airport Risk Propagation in Chinese Air Transport Network

Xuejun Zhang ^{1,2,3}, Shuaizhe Zhao,^{1,2,3} and Hao Mei^{1,2,3}

¹National Key Laboratory of CNS/ATM, School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

²Beijing Key Laboratory for Network-based Cooperative Air Traffic Management, Beihang University, Beijing 100191, China

³Beijing Laboratory for General Aviation Technology, Beihang University, Beijing 100191, China

Correspondence should be addressed to Xuejun Zhang; zhxj@buaa.edu.cn

Received 10 March 2021; Revised 2 January 2022; Accepted 21 January 2022; Published 1 March 2022

Academic Editor: Jing Zhao

Copyright © 2022 Xuejun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, due to the close coupling between airports, airport risk propagation has become a huge challenge. However, it has not been fully understood on the network level. Airport risk can be transferred through other airports owing to connected resources. In this study, we consider two risk factors including airport delay and saturation and propose a risk coupling model based on a clustering algorithm to fit the index and form risk series. To understand the risk propagation mechanism, we build risk propagation networks based on the Granger Causality test, and we apply complex network theory to analyze the evolution of the risk propagation network. We study the regular pattern of risk propagation from perspectives of time and space. Through network analysis, we find four time stages in the risk propagation process and the participation of airports in risk propagation has a positive correlation with airport sizes. In addition, more large airports tend to prevent risk propagation in unoccupied and normal situations, while small airports perform better than large airports in busy situations. Via the conclusion, our work can assist airlines or air traffic managers in controlling the scale of risk propagation before its key time turning point. By identifying the critical airport level and related factors in risk propagation, they can also reduce single airport risk and risk participation through corresponding risk control measures, finally avoiding the large-scale spread of risk and reducing delay or cancellation of more flights.

1. Introduction

With the rapid development of economic globalization, the world's civil aviation industry has grown fast [1]. The scale of China's civil aviation transportation industry is expanding sharply in the meantime. In the past 10 years, China's total air transport turnover has increased by 300%. Nevertheless, the total mileage of China's domestic air routes has only increased by 30%, which leads to increasingly intensified conflicts between limited air transport network resources and the rapid development of the civil aviation industry. The conflicts will cause the appearance of airport risk and aggravate the airport risk propagation. In this paper, the Air transport network (ATN) refers to the airport network. According to [2, 3], the airport network takes airports as vertices and direct flights between airports as edges. Like

above, the Chinese air transport network (CATN) refers to the Chinese airport network (CAN); in this paper, the edges of the Chinese airport network means there is a direct flight between two airports which is constrained by the prescribed airline, and the two airports will be added to the network node set, so the daily structure of CAN will change with the flights in China on that day. And the risk in this paper means the poor operation of the airport, for example, the airport operation efficiency is reduced and the turnover is difficult. The risk is defined considering two risk factors including airport delay and congestion in this paper.

In a complex network, the occurrence of node risk will cause the load of the node to be transferred to other nodes, which will lead to the paralysis of the entire network [4]. Similarly, ATN involves a large number of flights, complex airline structures, uncertain disturbances, and other factors.

Uncertain disturbance factors can be transferred through ATN among airports, which will cause widespread ripple effects. In ATN, the primary risk of an airport results from various disturbances, including inclement weather conditions, flight delays, huge flow, air traffic accidents. Similar to flight delay [5], airport risk in ATN can also be transferred and be amplified through other airports and uncertain disturbance factors because of connected resources.

For example, on July 21, 2012, the rainstorm in Beijing had a great impact on the safety of a wide range of flights across the country. The traffic capacity of PEK airport dropped by 50%, causing thousands of flights to be canceled or delayed. As a result, the risk of PEK propagates to other airports and contributes to their risks. It further leads to the decline of the overall operation efficiency, capacity loss, large-scale flight delay, or cancellation of the aviation network, and finally resulting in huge economic losses or the decline of airline reputation. Therefore, airport risk propagation in CATN has become a huge challenge and also needs attention. Accordingly, if we study the risk propagation regular patterns of CATN, such as the spatial and temporal characteristics of risk propagation, we can effectively control the scale of risk propagation and reduce the global network risk caused by a local risk which can finally reduce losses.

Therefore, in this study, we built a risk propagation network (RPN) based on the Granger causality (GC) test and investigated the direction and range of risk propagation at the network-level. We try to analyze the law of RPN from the perspectives of time and space. From the time angle, we use the size of the largest connected component of RPN to analyze the evolution process of risk propagation scale with time, try to reveal the time characteristics of its diffusion and dissipation in a day, and get its key time turning point. The results reveal that there were 4 stages of risk propagation of a day: dissipation of remaining risks, generation, fluctuation, and dissipation of intraday risks. We should pay attention to several key periods to better control risk propagation. From the spatial angle, we analyze the role of airports of different categories in risk propagation and identify the key airport category and related factors in the process of risk diffusion and absorption. The results show that the participation of airports had a positive correlation with their sizes and number of flights, and participation of large airports is higher than middle airports and small airports, so large airports are more noteworthy in risk propagation control.

Our results provide support to air traffic managers (ATM) on decision-making. According to the findings, the air traffic managers can develop effective countermeasures to prevent the risk propagation in particular links before key time points to alleviate network-wide flight delay or congestion and finally reduce the global risk of ATN. Moreover, through the causality test, if one variable acts as the cause for another one, actively intervening on the first would lead to the changes in the second. Thus, policy makers could potentially adopt the proposed method to analyze the interaction of airports in order to identify critical ones. It will help them to make decisions on resource allocation for improving airport capacity. They can better control risk propagation in

the future according to the key periods in a day, high participation airports, and vital flow range in our analysis.

Beyond this introduction, this work is organized as follows. Section 2 introduces some relevant previous studies and the differences between this work and previous work. Section 3 presents the materials and methods here applied, including the risk coupling model (Section 3.1), a description of Granger Causality test (Section 3.2), and some topology parameters of network analysis (Section 3.3). Section 4 introduces the construction of RPNs and analysis of the mechanism of risk propagation in CATN. Finally, Section 5 summarizes our major conclusions.

2. Literature Review

In recent years, researchers around the world have largely investigated the risk assessment of the civil aviation industry [6] as well as the propagation of single risk factors such as flight delay [7–11], congestion [12–14], and aircraft conflicts [15]. Although there has been rich research on assessment and propagation of risk in the civil aviation domain, they mainly focused on the civil aviation macroindustry with less consideration on the network level.

Complex network theory has achieved mature results in many fields such as identification and evaluation of influential nodes [16, 17], network risk assessment [18, 19], and decision-making based on the safety risk suggestions [20, 21] in recent years. These studies have important theoretical value for mining risk propagation mechanisms and safety risk assessment. Specifically, complex network theory is a powerful tool in the study of transportation systems [22, 23], with many network models defined for air transport system [24].

There is also some research that analyzes flight delay propagation on the network level; [25] developed the maximum connected subgraph of congested airports for assessing the level of delays across the entire system. Afterward, via establishing a delay propagation network based on a causal model, [26] used two models to simulate flight delay propagation and assessed the effect of disruptions in the US and European aviation networks. Reference [27] created a multilayer delay propagation network of 50 busiest airports and 20 airlines in Europe in which each airline acts as a layer for comparing the differences between 20 single layer networks and 3 projection networks. Reference [28] created a delay causality network based on the Granger causality test to understand the mechanism of flight delay propagation at the system level. Reference [29] built a delay propagation network based on the Bayesian Network approach to study the complex phenomenon of delay propagation within the network consisting of the 100 busiest airports in the United States. Reference [30] proposed an adaptation of machine learning's clustering analysis to air transport delay propagation network. By using Granger tests, they can assign each airport to the corresponding cluster, which includes mostly forcing node, intermediaries node, and mostly forced node. To the best of our knowledge, research on risk propagation mechanisms at the level of ATN is timely yet challenging and has not attracted much attention. In addition, the above previous studies mainly focused on single local risk factors

such as delay and congestion without full consideration of overall airport risk. There is also no research on comprehensive risk propagation.

Different from previous studies, we investigate the overall airport risk propagation with consideration of combining different local risk factors. Besides, applying complex network theory, we study the mechanism of risk propagation on the network level instead of civil aviation macro industry or local areas. In this paper, we propose the use of complex network theory to understand the mechanism of risk propagation in CATN with empirical analysis. For this purpose, we firstly propose a risk coupling model to obtain airport risk time series through multiple local risk factors. Then, we apply a risk propagation identification model based on the well-known Granger Causality (GC) test [31, 32] due to its primary advance on the relevance problem [33]. GC is a powerful tool to determine if two time series are related. Finally, we construct a risk propagation network (RPN) based on GC for each hour of a day, then use complex network theory to investigate how airport risks appear, propagate, and dissipate in CATN with empirical analysis.

3. Materials and Methods

3.1. Risk Coupling Model. Comparing with studies on simple risk factor ranking, we consider the impact of different risk factors and form risk coefficients. Instead of a rough assessment that only gives the risk level, we provide specific numeric results of risk, which is more intuitive and convincing. According to the flight statistics of the Civil Aviation Administration of China (CAAC), the number of air traffic accidents has increased and the rate of regular flights has decreased in recent years, which results in the risk of security and efficiency of CATN operation. Therefore, considering temporal and spatial properties, we select two local risk factors to quantify airport risk. In this paper, the risk means the poor operation of the airport, for example, the airport operation efficiency is reduced, and the turnover is difficult. The risk in this paper is obtained by coupling the delay weight and congestion weight. The delay factor reflects airport efficiency from a temporal perspective and the congestion factor reflects airport security from a spatial perspective. These two risk factors can reflect the operation efficiency change and safety situation of airports and reveal its comprehensive risk.

In recent years, average departure delay has also attracted much attention. Research on departure delay optimization [34, 35] and prediction [36, 37] has achieved prominent results. For temporal local risk factors, we adopt average departure delay due to its significant position in air traffic risk assessment. In addition, the average departure delay of airports represents the detention of flights and focuses on measuring the operation efficiency of CATN in the time dimension. We focus on hourly time series as we intend to investigate changes in risk propagation of a day with time series of each hour. For airport i , we construct its departure delay time series D_i by splitting one hour into 12 time intervals. Each time interval contains 5 minutes. The

value of each time interval represents the average departure delay. The average departure delay for airport i of time interval t is defined as follows [28]:

$$d_i(t) = \frac{y_i(t) + c_i(t) * b}{n_i(t)}, \quad t \in \{0, 1, 2 \dots 11\}, \quad (1)$$

where $y_i(t)$ represents the total departure delay of airport i during $(t, t + 1)$. $c_i(t)$ and $n_i(t)$ represent the number of cancelled flight and the total number of flights including canceled flights and scheduled flights during $[t, t + 1]$. Cancellation of flights should be taken into consideration in delay assessment [38]. b is the equivalent delay of canceled flights and generally equals 3 hours.

As for spatial local risk factors, congestion is a representative index in ATN and has attracted more and more attention in recent years [12–14]. The airport saturation describes the load situation of airports and focuses on the operation safety of ATN in the spatial dimension. As a result, we utilize airport saturation as a spatial factor to describe the congestion in CATN. For airport i , we construct its saturation time series S_i by splitting one hour into 12 time intervals. Each time interval contains 5 minutes. The value of each time interval represents the saturation. The saturation of airport i of time interval t is defined as follows:

$$s_i(t) = \frac{Q_i(t)}{C_i(t)}, \quad t \in \{0, 1, 2 \dots 11\}, \quad (2)$$

where $Q_i(t)$ represents the total arrival flights number of airport i during $[t, t + 1]$. $C_i(t)$ represents the capacity of airport i . In our work, airport capacity refers to the number of arrival aircraft that the airport can handle in an hour. With this method, we can construct delay and saturation as hourly time series.

After modeling risk factors, the next work is to couple them into the overall risk time series. For risk factor coupling, Li [14] utilized a grey clustering model to divide several local risk factors into 4 levels and weight them with a linear mapping function to obtain the final risk level. However, specific risk quantification results were not given. We propose a risk coupling model based on a clustering algorithm to obtain the overall risk coefficient. The specific method is defined as follows.

Firstly, we utilize (1) and (2) to obtain delay time series and saturation time series of each hour of a day. As delay and saturation have different dimensions; for both of them, we normalize each value of time series with the hyperbolic tangent function defined as follows:

$$d'_i(t) = \frac{e^{d_i(t)} - e^{-d_i(t)}}{e^{d_i(t)} + e^{-d_i(t)}}, \quad (3)$$

$$s'_i(t) = \frac{e^{s_i(t)} - e^{-s_i(t)}}{e^{s_i(t)} + e^{-s_i(t)}},$$

where $d'_i(t)$ is the normalized value of delay and $s'_i(t)$ is the normalized value of saturation of airport i in time interval t .

Secondly, in order to take advantage of internal relevance among airports and flights, we apply a K-means clustering algorithm [39] for delays of all airports in the same time

interval of an hour of a day and obtain 4 delay intervals. The same as delay, we adopt the same method to saturation and obtain 4 saturation intervals.

Then, according to the corresponding delay interval and saturation interval of airport i , we transform delay and saturation of airport i into delay weight $W_d(i, t)$ and saturation weight $W_s(i, t)$. We refer to the radial basis function in machine learning to define the delay weight mapping function and saturation weight mapping function in this paper, as shown in the following equations:

$$W_d(i, t) = \lambda_{d,c}^{0.5} * e^{d_i(t) - \lambda_{d,c} / \lambda_{d,c}} + \sum_{j=1}^{c-1} \lambda_{d,j}^{0.5}, \quad (4)$$

$$W_s(i, t) = \lambda_{s,c}^{0.5} * e^{s_i(t) - \lambda_{s,c} / \lambda_{s,c}} + \sum_{j=1}^{c-1} \lambda_{s,j}^{0.5}, \quad (5)$$

$W_d(i, t)$ and $W_s(i, t)$ corresponds, respectively, to the delay weight and saturation weight of airport i in time interval $[t, t+1]$. $W_d(i, t)$, c is the index of delay and saturation interval. $\lambda_{d,c}$ and $\lambda_{s,c}$ correspond, respectively, to the right boundary of delay interval c and saturation interval c , on the left of (4) and (5), the value of c depends on which cluster interval $d_i'(t)$ and $s_i'(t)$ is in. $d_i'(t)$ and $s_i'(t)$ represent, respectively, the normalized delay and saturation of airport i in time interval t . n_s is the number of intervals and n_s in our paper and $c \in \{1, 2, 3, 4\}$, which represents the first to fourth intervals.

Finally, we obtain the overall risk coefficient $R_i(t)$ airport i during $[t, t+1]$ according to the following equation:

$$R_i(t) = \sqrt{W_d(i, t)^2 + W_s(i, t)^2}. \quad (6)$$

As a result, we apply this model to airport k for each time interval to obtain risk time series (R_i) in an hour. Then, we calculate all risk time series of 24 hours in a day for all airports with this method. In this way, we have 24 groups of risk time series. Each group has the same number of risk time series as the number of airports.

3.2. Risk Propagation Identification Model. For risk propagation identification, we adopt the well-known Granger Causality (GC) test to test the causality between risk time series of 2 airports and identify the risk propagation according to the test result. GC test is proposed by the economy Nobel Prize winner Clive Granger. It is a statistical hypothesis test for determining whether one-time series is useful in forecasting another. For risk propagation, if the risks observed at one airport can explain the risks appearing at a second one, there exists a causality relationship, and we can identify the risk propagation from this airport to the second airport. Comparing with traditional methods, GC test is independent of airlines and air routes. GC test can identify risk propagation relation between any airports, ignoring their airlines and air routes. As a result, GC test can discover some undirect risk propagation relationships. In addition, GC test reveals propagation relationship from the perspective of causality, which is more profound.

GC test is sensitive to the stationarity of time series. As a result, we here apply a Z-Score detrend procedure to reduce the nonstationarity of the time series, which can result in a biased evaluation of the Granger Causality metric. In detail, the detrended risk time series for one airport is calculated as follows:

$$R'(h, t) = \frac{R(h, t) - R(., t)}{\sigma(R(., t))}. \quad (7)$$

$R'(h, t)$ represents the detrended risk for hour h at time interval t . $R(h, t)$ is the original risk. $\langle R(., t) \rangle$ and $\sigma(R(., t))$ are the average and standard deviation of the risk time series, respectively. Then, an augmented Dickey-Fuller test (ADF) [40] is applied to verify whether the risk time series is stationary. The risk time series difference procedure is applied until the time series is stationary. The time series difference is presented as follows:

$$R'(h, t) = R(h, t+1) - R(h, t), \quad (8)$$

where $R(h, t)$ and $R(h, t+1)$ are the risk for hour h at time interval $[t, t+1]$, respectively. And $R'(h, t)$ is the risk after difference procedures.

In our research, causality reveals the impacts between airport pairs and reflects the interaction of airport risks. Here, GC will help understand the existence and direction of the risk propagation between two airports based on the risk time series. For a pair of airports, an airport j “Granger-causes” another airport i if the use of past values of the risk time series of j helps improving the prediction of risk time series of i . In mathematical terms, suppose that the risks of two airports i and j can be expressed as two stationary time series Y_i and Y_j . Then, Y_j “Granger-causes” Y_i if the following equation can be satisfied.

$$\sigma^2(Y_i|U^-) < \sigma^2\left(\frac{Y_i|U^-}{Y_j}\right). \quad (9)$$

$\sigma^2(Y_i|U^-)$ is the square of residual in forecasting the time series Y_i using the past information of the entire universe U , and $\sigma^2(Y_i|U^-/Y_j)$ stands for the square of residual when the information about time series Y_j is discarded. The specific process of the GC test is presented as follows.

Firstly, the Granger causality test adopts an unrestricted regression equation to obtain its residual sum of squares RSSr in the following equation:

$$y_i(t) = \sum_{m=1}^P a_m * y_i(t-m) + \sum_{m=1}^P b_m * y_j(t-m) + e(t). \quad (10)$$

$y_i(t)$ is the current value of time series Y_i , $y_i(t-m)$ is the past value of time series Y_j , $e(t)$ is the error term, and a_m and b_m are the regression coefficients. In addition, P stands for the lag, indicating that the current value should be regressed with the past P values. Statistical analysis shows that the average flying time between all airport pairs in CAN is around two hours; thus we adopt the lag value $P = 2$ in this paper. Then, the null hypothesis that j does not cause i is defined as follows:

$$b_1 = b_2 = \dots = b_p = 0. \quad (11)$$

Secondly, we apply a restricted regression equation to obtain its residual sum of squares RSS_{ur} in the following equation:

$$y_i(t) = \sum_{m=1}^P a_m * y_i(t-m) + e(t). \quad (12)$$

Finally, we apply F-test to the model. F-statistic and p -value are adopted to test the null hypothesis as follows:

$$F = \frac{(RSS_r - RSS_{ur})/P}{RSS_{ur}/(n-P)}. \quad (13)$$

where n is the sample size of each time series. When the p -value is less than the chosen significance level ∂ (5% by default) of F-test and $>F_{\partial}$, it can be considered to pass F-test. In addition, we here apply the likelihood-ratio (LR) test [41] to the null hypothesis owing to insufficient length of risk time series. As a short length of time series may cause a fake causality relationship, LR can test the fitness of the model from the perspective of likelihood to increase accuracy of causality test. LR test is adopted as follows:

$$G = 2 * (\ln(L_{ur}) - \ln(L_r)), \quad (14)$$

where L_{ur} is the maximum likelihood value of the unconstrained regression model, and L_r is the maximum likelihood value of the constrained model. G is the test result of LR and approximately obeys the chi-square (χ^2) distribution. When the p -value is less than the chosen significance level ∂ (5% by default) of the chi-square test and $G > G_{\partial}$, it can be considered to pass the LR test.

When it passes both of F-test and LR test, the null hypothesis is rejected. If item Y_j belongs to this regression, Y_i is the cause Y_j . As a result, the value in j is partly attributed to i and we can consider that the risk of airport j propagates to airport i . In RPN, there is a directed link from airport j to airport i . For each pair of the airport which passed GC test, the link between them has no specific weight, and all the weight defaults to 1.

3.3. Network Analysis Metrics. Due to a large number of airports and complex interactions, the features of risk propagation cannot be understood from the information at the individual airport level alone. Complex network theory and its associated metrics and tools present an opposite approach to study ATN beyond what is offered by classical techniques [42]. Thus, a network-level analysis is adopted to capture the characteristics of risk propagation in CATN. Some topology parameters with practical significance are introduced here to help analyze network-level risk propagation.

It is specified that $A_{i,j} = (a_{i,j})_{N \times N}$ is the adjacency matrix of the network, where N is the number of network nodes. $a_{i,j} = 1$ if there is a directed edge from node i to node j ; otherwise, $a_{i,j} = 0$. The number of nodes is the number of individual points in the network, and the number of edges is

the number of links between nodes. In the undirected network, $a_{i,j} = a_{j,i}$.

- (1) Degree of node corresponds to the number of links with it. The node i has two types of out-degree (k_i^{out}) and in-degree (k_i^{in}) links, representing the number of nodes pointing or pointed by this node. The mathematical form is listed as follows:

$$k_i^{\text{out}} = \sum_{j=1}^N a_{i,j}, \quad (15)$$

$$k_i^{\text{in}} = \sum_{j=1}^N a_{j,i}. \quad (16)$$

- (2) Largest connected component [43] (LCC) is the largest group in which nodes are connected by existing edges with each other. It is generally introduced to reflect the scale of the network. In our research, we can assess the extent and seriousness of risk propagation through the size of LCC (N_{LCC}), which denotes the number of nodes in LCC.
- (3) Clustering coefficient (CD) [44] is used to qualify the inherent cluster tendency of nodes. The clustering coefficient of a node is the fraction of pairs of its neighbor nodes that have a link (i.e., the number of triangles in the network). For a network, the overall clustering coefficient is calculated as follows:

$$C_D = \frac{1}{2N} \sum_{i=1}^N \frac{\sum_j \sum_h (a_{i,j} + a_{j,i})(a_{i,h} + a_{h,i})(a_{h,j} + a_{j,h})}{d_{\text{single}} * (d_{\text{single}} - 1) - 2 * d_{\text{double}}}, \quad (17)$$

where $d_{\text{single}} = \sum_{j \neq i} a_{i,j} + \sum_{j \neq i} a_{j,i}$ and $d_{\text{double}} = \sum_{j \neq i} a_{i,j} * a_{j,i}$.

- (4) Reciprocity (RC) [45, 46] of the network reflects the bidirectional nature of links between airport pairs. The reciprocity means that node i affects airport j , whereas airport j also affects airport i . It is defined as follows:

$$R_C = \frac{\sum_{i \neq j} (a_{i,j} - \bar{a})(a_{j,i} - \bar{a})}{\sum_{i \neq j} (a_{i,j} - \bar{a})^2}, \quad (18)$$

where $\bar{a} = \sum_{i \neq j} a_{i,j} / (N(N-1))$ and n is the number of nodes.

- (5) Link density (LD) [47] describes the density of edges in the network. In the risk propagation network, it describes the density of risk propagation links between airports. The link density is relative to the number of nodes. It is defined as follows:

$$L_D = \frac{\sum_{i,j} a_{i,j}}{N(N-1)}, \quad (19)$$

where $a_{i,j}$ is the existing links in the network and N is the total number of nodes.

- (6) Efficiency (E) [48] of a network represents the ease for one node to reach other nodes (i.e., how many nodes have to pass by in order to reach the destination). The efficiency is defined by considering the inverse of the harmonic mean of the distances between pairs of nodes as follows:

$$E = \frac{1}{N(N-1)} \sum_{i,j \neq i} \frac{1}{d_{i,j}}, \quad (20)$$

where $d_{i,j}$ stands for the distance between node i and node j .

4. Results and Discussions

The flight data analyzed in this paper involves all domestic flight information in February, March, and May 2019 in China. The data of 3 months contained a total of 795973 domestic scheduled flights connecting 223 airports. The average number of daily flights is 8844. As a traditional Chinese holiday, May 1, 2019, was the busiest day with 9421 domestic flights. In contrast, as the Spring Festival of 2019, February 5 was the most unoccupied day with 6952 domestic flights. March 2, 2019, was a normal day with 8886 flights which is close to the average level. The weather is known to be one of the major factors influencing flight delays. By querying the air traffic website of the Civil Aviation Administration of China (CAAC) that publishes flight delay information, we can know the weather condition for the selected three days. On May 1, 2019 and March 2, 2019, there was no flight delay and airport capacity decline caused by weather. On February 5, 2019, only Wuhan Tianhe International Airport(WUH) was affected by the low visibility weather, the capacity of WUH was reduced by 30%, and other airports in CAN were not affected by the weather. Therefore, the selected three days in this paper can basically ignore the flight delay caused by weather factors and more effectively analyze the risk propagation of CATN under different dates. Details of the data set are shown in Table 1.

In order to perform a network-level analysis of risk propagation, we build an RPN using the pairwise GC test based on the flight data described in Table 1 for each hour of a day. For process and temporal characteristics of risk propagation, we focus on the hourly time series on March 2, 2019, due to its normal level of flight flow. For spatial characteristics, we focus on the features of airports of different sizes and their behaviors in risk propagation. We adopt the hourly time series on February 5, 2019, March 2, 2019, and May 1, 2019, to compare the features of different airports in different situations.

4.1. Process of Risk Propagation. The propagation process is a significant item to consider as it intuitively reveals how risk propagates in CATN. Among all parameters of a complex network, the size of LCC (N_{LCC}) can reflect the extent of risk propagation. We adopt N_{LCC} as it can directly help understand the generation, fluctuation, and dissipation of risk propagation during the propagation process. In order to

TABLE 1: Available information for each flight.

Information for each flight	
Airline	Flight code
Scheduled departure time	Actual departure time
Scheduled arrival time	Actual arrival time
Flight departure airport	Flight state
Flight arrival airport	Flight date

reveal the general situation of the risk propagation process, we focus on the flight data on March 2, 2019, and construct an RPN for each hour of 24 hours. There are a total of 223 airports. For one RPN of each hour, 49729 (223×223) times GC tests are performed. After removing airports with no connections, we generate 6 snapshots of RPN of different hours on March 2, 2019, in Figure 1 to present an obvious and visualized process of risk propagation.

During the process of risk propagation in a day, risk propagation exhibits cyclical fluctuations due to daily schedules. We investigate the appearance, fluctuation, and dissipation of risks in the process of risk transmission by observing the changes of NLCC. There are 4 stages in the process of propagation shown in Figure 2 of a day: dissipation of remaining risks, generation, fluctuation, and dissipation of intraday risks. In the stage of dissipation of the remaining risks of yesterday, the impact of risks of yesterday has not disappeared. The extent of risk propagation decreases until 5:00 with fewer scheduled flights. In the generation stage, the extent of risk propagation increases rapidly owing to the sharp growth of scheduled flights. Most busy airports generate initial delays due to the “Morning Rush” phenomenon. In the fluctuation stage, through the connection of resources, risk propagation of airports becomes serious with the increasing scheduled flights. In addition, the scale of risk propagation fluctuates at a high level. In the stage of dissipation of intraday risks, the extent of risk propagation decreases smoothly. Risk propagation will disappear the next day. The Air transport network will recover from the risk propagation due to the removal of flights from the system. Therefore, we need to pay attention to several periods as 5:00–6:00, 9:00–10:00 and 20:00–21:00, which are the key periods of risk propagation in a day. By improving the risks in these periods, we can better control risk propagation in a day.

In addition, we compared the evolutionary relationship between NLCC and delay, flow, and risk, respectively, as in Figure 3. Delay corresponds to the hourly average departure delay on March 2, 2019, and flow stands for the total number of scheduled flights for each hour on March 2. Risk represents the hourly average risk coefficient of airports which we obtained in Section 2.1. As Figure 3(b) shows, NLCC has a strong correlation with the flow. They have a similar evolutionary trend. However, as Figure 3(a) shows, there are some deviations between NLCC and delay due to inertia of delay. Flow can decrease in time with the decreasing scheduled flights. However, delay can accumulate gradually with the increasing flights and last for a long period, which is the “Cumulative Effect” of delay. Although the number of flights decreases, it still needs time to recover and decrease, which is the reason that delay continues to maintain at a high

FIGURE 1: Snapshots of RPN on March 2, 2019. (a) 03:00. (b) 07:00. (c) 11:00. (d) 15:00. (e) 19:00. (f) 23:00.

FIGURE 2: Stages of risk propagation process.

level at the end of the day. Risk is influenced by delay and flow and it is the result of them. As a result, risk combines their characteristics and has a similar evolutionary trend with NLCC, as Figure 3(c) shows. Results also reveal the close relationship between risk factors and NLCC.

4.2. Temporal Features. In addition to the process of risk propagation, we will investigate the evolution process of RPN to understand the temporal properties of risk

propagation. We adopt clustering coefficient, reciprocity, link density and efficiency of the network to analyze the evolution characteristics of RPN.

As results are shown as Figure 4, network clustering, reciprocity and link density have a similar evolutionary trend as NLCC, which means that they have a strong correlation with each other. High link density indicates a large number of edges in RPN, which causes a high level of network clustering and efficiency. Particularly, they are all temporarily at a lower level in a short moment from 13:00 to 14:00 due to fewer flights after morning rush hours. Nevertheless, reciprocity is special, as Figure 4(b) shows. The low reciprocity in the generation and fluctuation stage reveals that the single-direction propagation plays a significant role. In contrast, reciprocity sharply increases at 2:00 owing to the little number of flights and the significant role of double-direction links. Since other network metrics are more sensitive to changes in the number of network nodes and edges, they have similar evolution processes. However, reciprocity is independent of the number of network nodes and edges and reveals the bidirectionality of the network. This is the reason why reciprocity is different from other metrics.

4.3. Spatial Features. The global features of risk propagation are analyzed by the RPNs' properties, while risk propagation cannot be fully understood without analyzing the spatial

FIGURE 3: (a) Evolutionary relationship between (N)LCC and delay. (b) Evolutionary relationship between (N)LCC and flow. (c) Evolutionary relationship between (N)LCC and risk.

features. We investigate the spatial features from the point of view of behaviors of airports with different sizes in risk propagation. We divide airports into three categories according to their average departure and arrival flow within a day in Table 2.

The “Amount” in Table 2 means airport amount and the “Flow” means the number of flights. To understand the behaviors of airports in risk propagation, we will analyze the airports’ participation in different categories and the impact of airports on risk propagation. We have three different categories of airports in this paper including “Large airports,” “Middle airports,” and “Small airports”. The collective participation $P_c(t)$ of category c of airports is defined as the percentage of the number of airports involved in risk propagation belonging to category c to the number of total airports in category c . It is shown as follows:

$$P_c(t) = \frac{n_c(t)}{N_c}, \quad (21)$$

where $n_c(t)$ is the number of airports involved in risk propagation belonging to category c during $[t, t + 1]$, and N_c is the total number of airports in category c .

We focus on the flight data on February 5, 2019, March 2, 2019, and May 1, 2019, to compare the participation of different categories in different situations. As Figure 5 shows, participation is related to the category of airports. Participation of large airports is the highest in all of the situations, which indicates that most large airports are involved in risk propagation regardless of situations. Although small airports represent the majority of the system, they are rarely involved in the spread of risks. In addition, the hourly average participation $\overline{P(t)}$ of different airports increases with increasing flights, as Table 3 shows.

In addition, we will also analyze the impact of airports on risk propagation. To explore how airports affect risk propagation, we define the absorption coefficient r_i to assess the absorption ability of airport i for risk propagation in

$$r_i = \frac{\overline{k_i^{\text{in}}} + 1}{\overline{k_i^{\text{out}}} + 1}, \quad (22)$$

where $\overline{k_i^{\text{out}}}$ ($\overline{k_i^{\text{in}}}$) is the hourly average out-degree (in-degree) of airport i in risk propagation network over an entire day. k_i^{in} is the number of airports, each of which partly results in

FIGURE 4: (a) Evolution of average clustering (16). (b) Evolution of network reciprocity (17). (c) Evolution of link density (18). (d) Evolution of network efficiency (19).

TABLE 2: Airport information.

Airport size	Amount	Flow (x)
Large	35	$x \geq 150$
Medium	78	$10 \leq x < 150$
Small	110	$x < 10$

risk in airport i , while k_i^{out} is the number of airports whose risk is partly caused by airport i . For airport i , $r_i > 1$ means that outward links of airport i are fewer than its inward links, indicating that airports whose risks affect airport i are more than those whose risks are partly transferred by airport i , while $r_i < 1$ is the opposite. Therefore, we consider that airports with $r_i > 1$ have absorption ability for risk propagation and can mitigate the propagation, while those with $r_i < 1$ have a diffusion effect and aggravate the propagation. To understand the impact of each category of airports, we focus on the situations mentioned above. To analyze the effect of each kind of airports in risk propagation, we calculate the percentage of absorption airports ($r_i > 1$) a to the

number of airports that participated in risk propagation in the same category as follows:

$$a = \frac{n(r_i > 1)}{n_p}, \quad (23)$$

where $n(r_i > 1)$ represents the number of absorption airports, n_p is the total number of airports participating in risk propagation in the same category. High percentage of absorption means a significant role of airports in the absorption of risk propagation.

The final results are shown as Table 4. For February 5, 2019, and March 2, 2019, the majority of small airports exhibit the poor ability and low percentage of absorption of risk propagation due to their lack of resources and high vulnerability. As a result, more small airports tend to propagate risks under the influence of random factors. In contrast, large airports play a significant role in propagation absorption with a high percentage of absorption, indicating that the majority of large airports are affected by many upstream airports, but they impact fewer downstream

FIGURE 5: (a) Participation of airports on February 5, 2019. (b) Participation of airports on March 2, 2019. (c) Participation of airports on May 1, 2019.

TABLE 3: Average participation of airports ($\overline{P(t)}$) in risk propagation.

Airport size	February 5, 2019	March 2, 2019	May 1, 2019
Large	0.656	0.662	0.705
Medium	0.368	0.420	0.433
Small	0.049	0.073	0.077
Flow	6452	8886	9421

airports. In our expectation, the large airports are located in developed cities and bear the pressure of heavy traffic, as well as its large volume of flights with downstream airports, it should have a great impact on the downstream airports, but the findings are contrary to expectations. It may be because they also have sufficient resources and high robustness to partly resist the impact of delays and flow in the meantime. Therefore, they perform best in unoccupied and normal situations. The percentage of medium airports is at a

TABLE 4: Percentage of absorption airports (a) for risk propagation.

Airport size	February 5, 2019	March 2, 2019	May 1, 2019
Large	0.571	0.629	0.429
Medium	0.506	0.474	0.423
Small	0.456	0.459	0.500
Flow	6452	8886	9421

medium level. Nevertheless, the situation was just the opposite on May 1, 2019. Small airports performed best, while the majority of large and medium airports tended to propagate risks. Due to a large number of flights and heavy air traffic pressure on May 1, 2019, large and medium airports bear most of the traffic pressure. Since large airports can partly resist the impact of delays and flow, why did they perform worse than small airports? This is mainly due to the overload of parts of large and medium airports. When a load of some airports is high enough to cause overload, the load

will be transferred to other airports according to the links in spite of their high robustness. As a result, small airports receive more risks transferred from large airports due to their overload, which is the reason why small airports performed best on May 1, 2019. It is worth noting that there is a critical flow between 8886 and 9421 where the situation will become opposite for large and small airports. To better control risk propagation, we need to pay attention to the flow between 8886 and 9421 and investigate the critical flow, which is meaningful for CATN risk control and safety improvement.

5. Conclusions

In this paper, we investigated the mechanism of risk propagation among airports in the Chinese air transport network (CATN) from a new perspective, i.e., constructing a risk propagation network (RPN) based on Granger Causality (GC) for each hour of a day and applying network analysis tools to reveal the macroscopic appearance of risk propagation. To realize this model, we firstly built a risk coupling model to obtain the risk time series of airports from delay and flow. Then, we built hourly RPNs based on risk time series and GC test.

We firstly investigated the process of risk propagation in a normal situation and analyzed the properties of the propagation process. There were 4 stages of risk propagation of a day: dissipation of remaining risks, generation, fluctuation, and dissipation of intraday risks. We should pay attention to several key periods as 5:00–6:00, 9:00–10:00 and 20:00–21:00 to better control risk propagation. Then, for temporal properties, the evolution of reciprocity is relatively special due to its independence of the number of network nodes and edges. Finally, we investigated the spatial properties of risk propagation. We found that the participation of airports had a positive correlation with their sizes and the number of flights. Generally, participation in large airports is more than 60%, while participation in small airports is less than 10%. The difference is significant. As a result, large airports are more noteworthy in risk propagation control. For absorption ability, i.e., the ability to prevent risk propagation, we calculated the percentage of absorption airports and found that the largest airports performed best on February 5, 2019, and March 2, 2019, because of their sufficient resources and high robustness to normal situations, while the opposite on May 1, 2019, due to the overload of some large airports causing more risk transferred to small airports. Therefore, it is necessary to focus on the situations where flow is between 8886 and 9421 to prevent the overload of large airports.

The results of the analysis provide a theoretical basis for the optimization and management of air transport network safety and efficiency. In addition, the results provide some useful suggestions for reducing air transport network risks and enhancing network load capacity in order to establish an advanced intelligent air traffic system. We can better control risk propagation in the future according to the key periods in a day, high participation airports, and vital flow range in our analysis. Nonetheless, the study of using RPN to study

airport risk propagation could be extended further. For example, we found the phenomenon of overload of airports. It is interesting to compare more situations to obtain the critical value of flow to cause an overload of large airports. According to critical flow, we can better control the risk propagation and optimize the overall risks of airports. In addition, it is also meaningful to study the relationship between the number of daily flights and the absorption percentage of airports in detail.

Data Availability

The flight data used in this paper are available at <http://www.variflight.com/en/>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 71731001).

References

- [1] P. Belobaba, A. Odoni, and C. Barnhart, *The Global Airline Industry*, John Wiley & Sons, Hoboken, NJ, USA, 2009.
- [2] K. Cai, J. Zhang, W. Du, and X. Cao, "Analysis of the Chinese air route network as a complex network," *Chinese Physics B*, vol. 21, no. 2, Article ID 028903, 2012.
- [3] X. Sun, V. Gollnick, and S. Wandelt, "Robustness analysis metrics for worldwide airport network: a comprehensive study," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 500–512, 2017.
- [4] A. E. Motter and Y. C. Lai, "Cascade-based attacks on complex networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 66, Article ID 065102, 2002.
- [5] N. Kafle and B. Zou, "Modeling flight delay propagation: a new analytical-econometric approach," *Transportation Research Part B: Methodological*, vol. 93, no. partA, pp. 520–542, 2016.
- [6] J. Li, Y. Chen, and P. He, "Research on risk monitoring and simulation of China's civil aviation industry," *China Safety Science Journal*, vol. 1920 pages, 2009.
- [7] N. Pyrgiotis, K. M. Malone, and A. Odoni, "Modelling delay propagation within an airport network," *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 60–75, 2013.
- [8] M. Zanin, S. Belkoura, and Y. Zhu, "Network analysis of Chinese air transport delay propagation," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 491–499, 2017.
- [9] S. Belkoura, J. M. Peña, and M. Zanin, "Beyond linear delay multipliers in air transport," *Journal of Advanced Transportation*, vol. 2017, Article ID 8139215, 11 pages, 2017.
- [10] C.-L. Wu and K. Law, "Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model," *Transportation Research Part E: Logistics and Transportation Review*, vol. 122, pp. 62–77, 2019.
- [11] B. Baspinar and E. Koyuncu, "A data-driven air transportation delay propagation model using epidemic process

- models," *International Journal of Aerospace Engineering*, vol. 2016, Article ID 4836260, 11 pages, 2016.
- [12] M. Bernardo, V. D. P. Servedi, and L. Vittorio, "Congestion transition in air traffic networks," *PLoS One*, vol. 10, no. 5, Article ID e0125546, 2015.
 - [13] X. Dai, M. Hu, and W. Tian, "Modeling congestion propagation in multistage schedule within an airport network," *Journal of Advanced Transportation*, vol. 2018, pp. 1–11, Article ID 6814348, 2018.
 - [14] S. Li, *Research on Identification and Prediction Methods of Air Traffic Congestion*, Tianjin University, Tianjin, China, 2013.
 - [15] M. Zhang, B. Liang, S. Wang, M. Perc, and W. Du, "Analysis of flight conflicts in the Chinese air route network," *Chaos, Solitons & Fractals*, vol. 112, pp. 97–102, 2018.
 - [16] B. Song, G. Jiang, Y. Song, and L. Xia, "Rapid identifying high-influence nodes in complex networks," *Chinese Physics B*, vol. 24, no. 10, Article ID 100101, 2015.
 - [17] X. Yan, Y. Cui, and S. Ni, "Identifying influential spreaders in complex networks based on entropy weight method and gravity law," *Chinese Physics B*, vol. 29, no. 4, Article ID 048902, 2020.
 - [18] T. Bjerga, T. Aven, and E. Zio, "Uncertainty treatment in risk analysis of complex systems: the cases of STAMP and FRAM," *Reliability Engineering & System Safety*, vol. 156, pp. 203–209, 2016.
 - [19] Z. Mohaghegh, R. Kazemi, and A. Mosleh, "Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: a hybrid technique formalization," *Reliability Engineering & System Safety*, vol. 94, no. 5, pp. 1000–1018, 2009.
 - [20] W. Li, L. Zhang, and W. Liang, "An Accident Causation Analysis and Taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives," *Safety Science*, vol. 92, pp. 94–103, 2017.
 - [21] A. Nguyen, P. Do, and A. Grall, "Multi-level predictive maintenance for multi-component systems," *Reliability Engineering & System Safety*, vol. 144, pp. 83–94, 2015.
 - [22] M. Zanin, X. Sun, and S. Wandelt, "Studying the topology of transportation systems through complex networks: handle with care," *Journal of Advanced Transportation*, Article ID 3156137, 2018.
 - [23] X. Sun, S. Wandelt, and F. Linke, "On the topology of air navigation route systems," *Proceedings of the Institution of Civil Engineers - Transport*, vol. 170, no. 1, pp. 46–59, 2017.
 - [24] M. Zanin and F. Lillo, "Modelling the air transport with complex networks: a short review," *The European Physical Journal - Special Topics*, vol. 215, no. 1, pp. 5–21, 2013.
 - [25] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Systemic delay propagation in the US airport network," *Scientific Reports*, vol. 3, no. 3, p. 1159, 2013.
 - [26] B. Campanelli, P. Fleurquin, A. Arranz et al., "Comparing the modeling of delay propagation in the US and European air traffic networks," *Journal of Air Transport Management*, vol. 56, pp. 12–18, 2016.
 - [27] M. Zanin, "Can we neglect the multi-layer structure of functional networks?" *Physica A: Statistical Mechanics and Its Applications*, vol. 430, pp. 184–192, 2015.
 - [28] W.-B. Du, M.-Y. Zhang, and Y. Zhang, "Delay causality network in air transport systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 466–476, 2018.
 - [29] Q. Li and R. Jing, "Characterization of delay propagation in the air traffic network," *Journal of Air Transport Management*, vol. 94, 2021.
 - [30] M. Zanin, "Simplifying functional network representation and interpretation through causality clustering," *Scientific Reports*, vol. 11, no. 1, 2021.
 - [31] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
 - [32] C. W. J. Granger, "Some recent development in a concept of causality," *Journal of Econometrics*, vol. 39, no. 1–2, pp. 199–211, 1988.
 - [33] M. R. Frank, N. Obradovich, and L. Sun, "Detecting reciprocity at a global scale," *Science Advances*, vol. 4, no. 1, Article ID eaao5348, 2018.
 - [34] A. Montlaur and L. Delgado, "Flight and passenger delay assignment optimization strategies," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 99–117, 2017.
 - [35] T. Bolić, L. Castelli, L. Corolli, and D. Rignonat, "Reducing ATFM delays through strategic flight planning," *Transportation Research Part E: Logistics and Transportation Review*, vol. 98, pp. 42–59, 2017.
 - [36] G. Gui, F. Liu, J. Sun, J. Yang, Z. Ziqi, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2019.
 - [37] M. Zanin, Y. Zhu, R. Yan, P. Dong, X. Sun, and S. Wandelt, "Characterization and prediction of air transport delays in China," *Applied Sciences*, vol. 10, no. 18, Article ID 6165, 2020.
 - [38] J. Xiong and M. Hansen, "Value of flight cancellation and cancellation decision modeling," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2106, no. 1, pp. 83–89, 2009.
 - [39] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
 - [40] W. Cheung and K. S. Lai, "Lag order and critical values of the augmented dickey-fuller test," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 277–280, 1995.
 - [41] W. Vandaele, "Wald, likelihood ratio, and Lagrange multiplier tests as an F test," *Economics Letters*, vol. 8, no. 4, pp. 361–365, 1981.
 - [42] A. Cook, H. A. P. Blom, F. Lillo, and R. N. Zanin, "Applying complexity science to air traffic management," *Journal of Air Transport Management*, vol. 42, pp. 149–158, 2015.
 - [43] A. Broder, R. Kumar, F. Maghoul et al., "Graph structure in the web," *Computer Networks*, vol. 32, no. 1–6, pp. 309–320, 2000.
 - [44] G. Fagiolo, "Clustering in complex directed networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 76, no. 2, Article ID 026107, 2007.
 - [45] L. da Rocha, "The International School for Advanced Studies (SISSA), find out more Structural evolution of the Brazilian airport network," *Journal of Statistical Mechanics: Theory and Experiment*, Article ID P04020, 2009.
 - [46] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

- [47] M. D. Laat, V. Lally, L. Lipponen, and R. J. Simons, "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: a role for Social Network Analysis," *International Journal of Computer-Supported Collaborative Learning*, vol. 2, no. 1, pp. 87–103, 2007.
- [48] L. Vito and M. Massimo, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, Article ID 198701, 2001.