Airline Analysis Dashboard - An Analysis of Airline On-time Performance and Airport Connectivity

He Xi Shi Xiaoyu Wang Miaoqiong

Abstract—As air travel becomes more and more popular and commonplace, customers have higher standard for airline performance. The flight delay and cancellation can have a significantly negative impact on customer satisfaction and thus influence the market share of airline companies. Apart from that, the plane crash and airline bombing attacks prompt the public to be more alerted towards the security issue. Therefore, to improve the on-time performance as well as keeping the public far from potential risks, we propose an airline analysis dashboard as a solution to investigate overall airline cancellation and delay as well as detailed causes respectively. On top of that, we will also analyse the airport connectivity which can fasten the decision-making process when faced with terrorism attack or infectious disease outbreak. In the dashboard, the route map, bubble chart, bar chart, radar chart and network graph are used.

Index Terms - Airline Analysis, Delay, Cancellation, Network Connectivity, Route Map, Bubble Chart, Bar Chart, Radar Chart and Network Graph

1. INTRODUCTION

Flying has been one of the best way for travelling because of its speed, comfort and efficiency. It opens an unlimited number of possibilities for travellers, allowing them to experience different cultures and explore our planet to the fullest. With the steady decrease in the traffic cost, the demand for air travel is increasing significantly.

The line chart in the figure 1 shows the number of airline passengers has been increasing since 1970 and the increase is especially more drastic after stepping into 21 centuries. In 2016, the number of airline passengers was 3.7 billion around the world. As flying becomes more popular and commonplace, air travel is not only a novelty but an irreplaceable necessity. The increasing demand comes with the higher customer standard. Hence, a good on-time performance is very important for airline companies to improve its customer satisfaction and the market share.

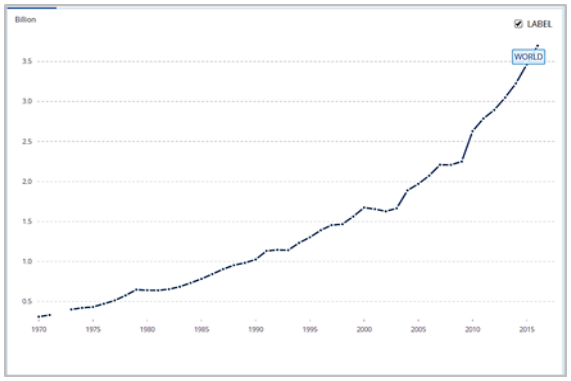

Figure 1 Line chart reveals number of world airline passengers from 1970 to 2016

However, despite all the advantages it has, the frequently occurring plane crash and airline bombing attacks expose its vulnerability to the public. It drives the public to have more strict requirements for aircraft security. Therefore, to protect the public from risks and meet the high requirements, airline companies should take actions to identify the potential threats and implement risk mitigation plans.

2. MOTIVATION AND OBJECTIVES

Flight delays has been a very common problem for travellers, the delay can be attributable to various problems, such as, aircraft issues, weather issues at origin airport or/and destination airport. The delay has no doubts will disappoint air travelers and affect their flight experience greatly. The passengers who experienced flight delays are more likely to switch airlines for the subsequent flight than those passengers who did not experience delays. Thus, in this project, our team aims to investigate the performance of different airlines and flight delays in detail.

Air transport network is a very critical and complex transportation infrastructure for a nation, it supports flow of passengers and business activities, such as logistics between different destinations and is increasingly important for public policy considerations. Failure in the air transport network, for example, caused by terrorist attack, disease transmission or other reasons, may result in severe disruptions for travellers and business activities which may eventually lead to significant time and economic losses. In our analysis, we applied social network analysis (SNA) methodology into our data to identify critical notes in air transport network, so as the users can take proactive measures to prevent occurrence of disruptions as well as to take effective mitigation plans to reduce the negative impact of the disruption.

As shown in the figure 2, the number of air traveler in US reaches 7.8 billion in 2015, the highest among all the countries around the world. Hence, our project decided to zoom into the US airline performance in 2015. The analysis can also be applied to other countries.

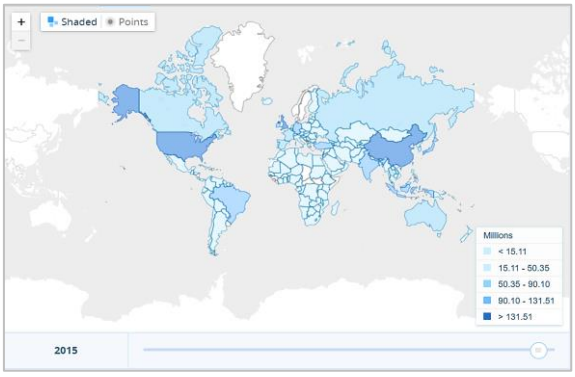

Figure 2 Heatmap reveals air traveler's density by country

3. VISUALIZATION DESIGN

3.1 Route Map for Airline On-Time Performance

Tableau is used to plot the route map which is effective in displaying geographical distribution based on a specific measurement. The chart shows an overview of the average arrival delay of each route for selected origin and destination airport. The darkness of the colour and bubble size represent the seriousness of the arrival delay. An interactive dropdown list is applied for users to view both incoming and outgoing flights for selected airport.

Figure 3 route map for visualizing on-time performance of flight route between different departure and destination airports

3.2 Bar Chart for Overall Delay by Airline

The two bar charts in are used to showcase the average delay and total delay of airlines in US in Jan 2015 respectively. The x-axis represents the arrival delay in minutes and the y-axis represents 14 airlines in US (UA, AA, UB, F9, B6, OO, AS, NK, WN, DL, EV, HA, MQ, VX). The positive value in x-axis indicates the flight is behind the schedule whereas the negative values shows the flight is ahead of the schedule.

The bar chart is chosen here because it's useful to display and compare the total/average values among different discrete categories (airline).

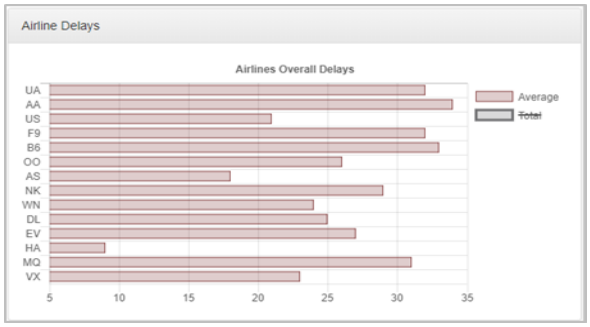

Figure 4 Bar chart for visualizing overall delay by airline

3.3 Radar Plot for Delay Causes by Airline

The radar plot is effective in showcasing the multivariate data in a two-dimensional chart of many quantitative variables represented on axes starting from the same point. The quantitative variables in the radar plot are five delay causes measured in minutes, which are air system delay, weather delay, late aircraft delay, security delay and airline delay. Drill down is also provided for different airlines. By selecting the airline, users can know the distribution of different delay causes.

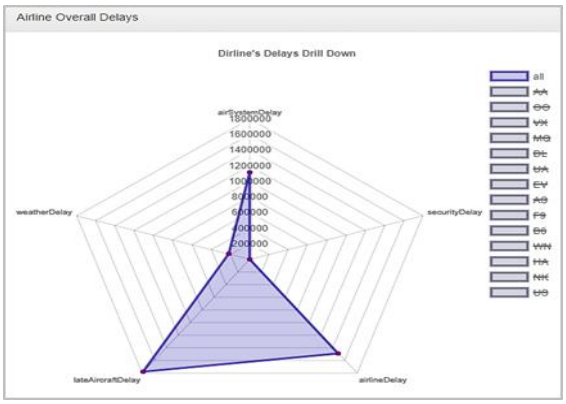

Figure 5 route map for visualizing time of delays breakdown by airline and delay reasons

3.4 Bubble Chart for Flight Cancellation

Bubble chart in D3.js is utilized to analyze the airline cancellation reasons. Due to processing capacity, in this chart, we only used partial records from January data. Each data point contains two attributes, airline and cancel reason. The goal of this visualization is to provide an overview of all cancelled airline. As there are various reasons (incl. A- Airline/Carrier, B- Weather, C- National Air System, D- security) for flight cancellation, an interactive design is applied to re-group the airline by different reasons of cancellation. By clicking on the blue tab highlighted in red box, readers can have a detailed view of distribution of cancellation reasons.

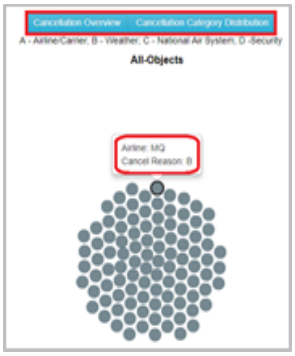

Figure 6 Bubble chart for overall cancellation by airline

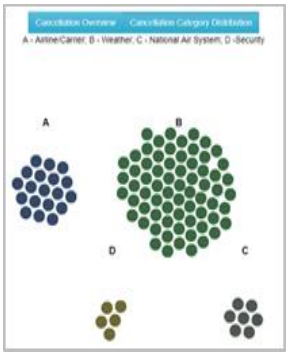

Figure 7 Bubble chart for cancellation breakdown by causes

3.5 Network Graph for Airport Connectivity

This graph is used for the demonstration of the U.S flight routes data as a connected network. With this graph, our team was trying to identify roles of different airports in the U.S air transport network by various centrality measures. Each node represents one airport and the arcs between nodes represent the flight routes. The opacity of colour shows the closeness centrality of each airport, while the size of nodes shows the betweenness centrality of each

Airport.

Betweenness centrality, the most powerful measure of centrality, is a measure of positional advantage such that indicates the number of times a node lies on shortest paths (geodesic) that connecting two other nodes together. In simple words, it counts the number of geodesic that pass through a node. Hence, the larger the betweenness centrality, the greater the potential to become a hub airport and transfer points due to its importance in connecting different airports. Furthermore, the airports with high betweenness centrality, governments should pay great attention to fortify from attacks when terrorist attacks happen.

Closeness centrality indicates how far a node (airport) is to all the other nodes in the network. It shows the extent of influence of a node on the network. Higher closeness centrality suggests closest between nodes while lower value suggests more distant. In terms of disease transmission, closeness centrality is used to measure vulnerability to disease.

Given that betweenness centrality and closeness centrality both take into account the shortest path, therefore, the increase of closeness centrality value will lead to increase of betweenness centrality.

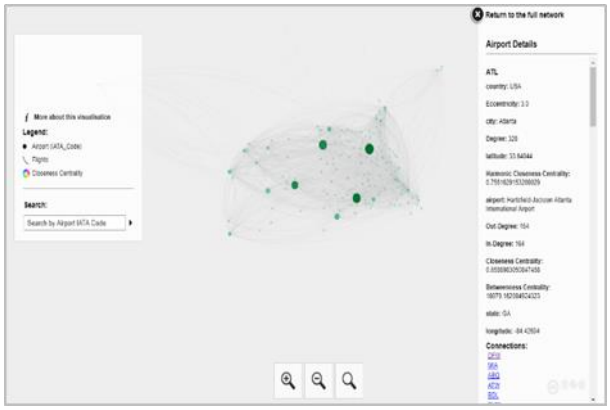

Figure 8 Network graph for visualizing different centrality measures

4. DESIGN OF APPLICATION

Our application consists of four parts, which are homepage for delay overview and problem statement, network graph for connectivity between airports, bar and radar plot for detailed delay and bubble chart for cancellation.

The home page starts with the problem statement and the motivations, followed by the route map of the airports and flights. The arrival delay is indicated by the darkness of the colour, from which users can have an overview of the on-time performance of all flight routes. In addition, it enables user to focus on specific flights by selecting origin airport and destination airport and detailed information tips appear when the mouse hovers on airports routes.

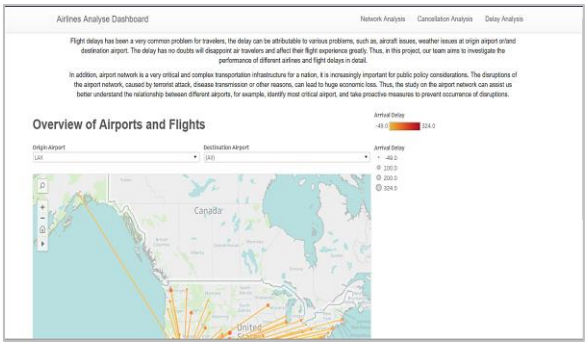

Figure 9 Home page

Users can navigate to the network graph by selecting the network analysis tab. When clicking on any node, a pane with airport details appear at right hand side, which contains all degree and centrality measurements. Apart from that, users are also allowed to search the airport in the search box at the left-hand side.

Figure 10 Network analysis page

Moving on to the cancellation analysis, users can have an overview of all cancellation flights by looking at the number of bubbles. In the cancellation category distribution, users can tell and compare the number of cancellation under each category.

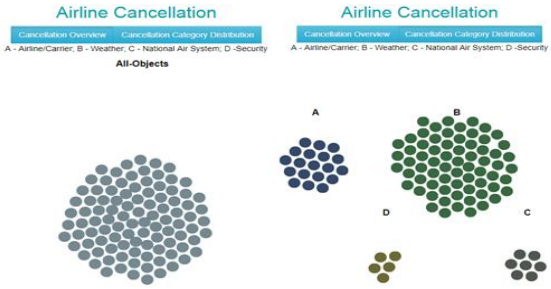

Figure 11 Cancellation analysis page

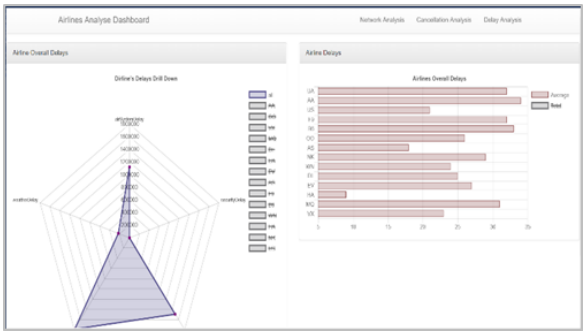

Figure 12 Delay analysis page

Last but not the least, there are one bar chart for overall airline delays and one radar chart for detailed delay causes for each airline under the delay analysis tab. From the bar chart, users can tell the average delay and total delay among all US airlines. The radar chart at the left displays the contribution of five delay causes to the overall delay for the airline.

5. DATA VISUALIZATION

5.1 Data Collection

The project begins with the selection of a topic and identification of problems. Then, we proceed to collection of data set that provides the ability to analyse airline on-time performance and analyse the connectivity of airports. The data source is taken from Kaggle which provides comprehensive information on the U.S domestic flights operated by large air carriers (e.g. United Air Lines Inc.) in 2015. This data contains a total of 5819080 records, due to the limited computing/processing capacity, we only used the January data which consist of about 470,000 records for our analysis.

5.2 Data Preparation

5.2.1 Understand the data

There are more than 30 columns in the source data while the metadata for the columns are limited and not clearly stated. To dig out more information about the data structure, we went through the department of transportation in US. We also conducted cross comparison among the columns to better understand the data.

5.2.2 Prepare required data in csv format

We selected the data that can best suit the project purpose and eventually only 18 columns are left. In addition, we transformed the tail_number by using the function of Mid (cell, 2, 3) to retrieve the aircraft type. We also replaced empty delay value with 0. Considering the limited computing and processing capacity, we separate the data into 12 parts by month. In this project, we use January data for analysis.

5.3 Data Visualization

In the homepage, we use HTML5 to embed the problem and motivation statement and tableau for the route map of arrival delays for all flights.

The Gephi is used to design the network graph to analyse the connectivity among airports. Gephi is very powerful tool which can create customized network graph with degree and centrality measurements. With sigma.js plug in, we can export the network graph into sigma.js file which can be easily integrated with our application.

Bubble chart and radar chart are implemented with HTML5 and D3. D3.js is a JavaScript library to construct dynamic, interactive data visualizations in web browser. As for the bar chart, it is designed in the chart.js. Chart.js is a JavaScript library which can create simple and customizable HTML5 based JavaScript charts.

6. RECOMMENDATION

From the visualization application, we would like to share our insights and offer some suggestions to US airline companies as well as the national security departments.

Let's first take a look at the overall delays (Radar plot), late aircraft delay is the most common delay, it happens when the previous flight with same aircraft arrived late, causing the present flight to depart late. From this chart, we can infer that it is necessary for US domestic flights to improve the flights scheduling and have better flight scheduling system in place to reduce late aircraft delay. Airline delay, is another major delay reason which is more controllable by air carrier and can be caused by various reasons, such as maintenance, aircraft damage, crew problems, fueling, etc.

Drilling down into each airline, we can observe that airline WN (WN, Southwest Airlines Co.) has longest delay time in January, and late aircraft delay and airline delay were the main reasons. Therefore, WN airline can also further investigate what is the root cause of airline delay. For example, if aircraft damage or maintenance is the major cause, proper maintenance and frequent inspections will reduce airline delay greatly.

From the network graph, we observe that airports, Dallas/Fort Worth International Airport (DFW), Hartsfield-Jackson Atlanta International Airport (ATL), Chicago O'Hare International Airport (ORD) and Denver International Airport (DEN) has highest betweenness centrality and closeness centrality value among other airports. Therefore, these four airports have greater potential to be a hub airport and transfer point in the U.S domestic air transport network. Due to its positional advantage, government should also strengthen the defences and enhance security of these airports to fortify from attacks when terrorist attacks take place. In addition, high closeness value also suggests that these four airports are particularly vulnerable to disease transmission when a highly infectious disease outbreak, therefore, they need to be closed immediately to effective quarantine the disease from spreading and reaching the rest of the airport.

7. CONCLUSION

The airline analysis dashboard is designed with the objectives of identifying areas of improvements on airline on-time performance and keeping the public far from potential risks by looking at the delay and cancellation causes in details and investigating airport connectivity. It is organized in a very simple and logical way. First of all, it provides an overview of on-time performance for all airlines. Then it zooms into cancellation and delay for each airlines respectively. Moreover, the visualization to detect which causes contribute most to the delay or cancellation are also provided. Last but not least, a comprehensive airport network is designed to show the connectivity among the airports. For this analysis, we only use one month data due to computing limitation but the same analysis can also be applied to other months. Furthermore,

the analysis can be expanded to other countries as well as other types of transportation.

REFERENCES

1. Borgatti, S. P. (2005). Centrality and network flow. Social Networks, 27(1), 55-71
2. Bureau of Transportation Statistics, https://www.rita.dot.gov/bts/ help/aviation/html/understanding.html
3. Danielatkin(2016), Interactive Bubble Chart, https://bl.ocks.org/danielatkin/57ea2f55b79ae686dfc7
4. Sophie Sparkeson (July 31, 2015), Taking off with the path function, https://public.tableau.com/en-us/s/blog/2015/07/taking-path-function
5. Min Geun Song et al. (2017), Analysis of the Air Transport Network Characteristics of Major Airports, http://www.sciencedirect.com/science/article/pii/S209252121730041X
6. David C Bell et al., Centrality measures for disease transmission networks, http://www.sciencedirect.com/science/article/pii/S0378873398000100