Efficiently Rerouting Flights in a Real-Time Air Traffic Simulation

Anthony Mainero, Tom Schmidt, Harley Sugarman

October 17, 2013

Abstract

The goal of our project is to build an algorithm for efficient and cost-effective rerouting of flights in an air traffic network when specific nodes (airports) become inaccessible. We will then test this algorithm by creating an air traffic simulation where nodes are randomly cut out from the graph at random times throughout the day, forcing re-routes for flights that have not yet left their departing airports and flights that are already in the air.

1 Related Work

1.1 Summary

1.1.1 Systemic Delay Propagation in the US Airport Network

In 2010, 37.5% of flights within the United States departed or arrived late, with an average delay of 29 minutes. In this paper, the authors model the US air travel network as a directed graph with a series of nodes (airports) and edges (flights), and simulate delays using three distinct metrics: aircraft rotation time (ART), flight connectivity (FC) and airport congestion (AC). Their initial findings indicated that delays followed an inverse-logistic curve, with most flights landing only a few minutes late. This curve was found to be particularly robust: the graphs were consistent regardless of the season, or whether the delay initiated at the airport of departure or arrival.

FC was found to be the most important factor in determining the length of delays. Decreasing the AC did not ease the propagation of delays since the main cause of the spreading, FC, is independent of it. Adding a buffer time to the ART did alleviate some delays, but not significantly enough to substantially modify the distribution.

1.1.2 Characterization of Delay Propagation in the US Air Transportation Network

The authors focus primarily on performing existing forms of analysis on a network composed of nodes representing all US airports and directed edges representing existing flight paths between these airports. They examine various graph metrics, including degree distribution and clustering coefficients. Looking at longest paths in the network reveals that a vast majority of US airplanes take between 2 to 7 flights per day. More interesting results come from looking at flight delay statistics. Flight delay times fall within an extremely wide range, peaking at around 700 minutes. When exploring the network on a node-by-node basis, the authors found that delays at more remote airports, such as Honolulu, caused a much wider range of problems than delays at larger hub nodes on the mainland. This was attributed to the longer duration of flights that were coming to or from Honolulu.

1.1.3 Immunization of Complex Networks

This paper discusses ways of preventing the propagation of viruses through a network (or in our case of air transportation, the propagation of delays). In order to maximize the effect of immunization, one should choose a few specific groups of localized key nodes to 'vaccinate', rather than treat the network from a global perspective.

1.2 Critique

1.2.1 Systemic Delay Propagation in the US Airport Network

Fleurquin, Ramasco and Eguiluz create a mathematical model of the US airport network and use real data to evaluate reasons for delay propagation. The analysis is sound, and their manipulation of different factors to determine the causes of delays is well thought-out. The authors also gain some useful insight into the rigidity of the delay network. However, the end result of their work is underwhelming in its attempt to produce a valid solution to the problem of air transportation delays. Rather than trying to envision a way to minimize airport delay propagation, they simply re-state that it is, in fact, a problem. The data is presented in an extremely static form, and practical applications of this research are often glossed over or ignored entirely.

1.2.2 Characterization of Delay Propagation in the US Air Transportation Network

While the authors' idea of modeling the US air transportation system as a network was a good one, and they were able to gain some valuable insights into the nature of delays, most of their analysis was very superficial and not specific to the domain of air transportation. For example, although they did an analysis of degree distribution and longest paths, they didn't explain why these stats were important in the context of air travel. Furthermore, the entire report is very passive. Instead of trying to generate new predictions or insights that could be applied to new data in the future, they simply reinforce and rehash old notions that are not particularly useful or exciting.

1.2.3 Immunization of Complex Networks

In the context of air transportation, the idea of selective group useful in generating a model to reduce the spread and severity of air travel delays. As long as certain 'hub' airports are able to reduce delays as much as possible, the butterfly effect of the delay 'virus' will be dampened throughout other airports in the network.

2 Proposal

2.1 The Plan: Construct a Flight Network and Simulate Airport Node Disconnections

We plan to build a simulation with predetermined data, populated from both real-world and generated sources, containing flights for a given day. The flight network is modeled as a list of planes, with each plane having a path (consisting of a list of edges between airports), as well as start and end times for each of these flights. At random times throughout the day, a random airport node will be removed from the network to emulate a delay or cancellation. All flight edges to that airport will be cut, and any planes involving that airport in the rest of their daily path must be re-routed.

When an airport node A is removed from the network, it creates a series of outgoing and incoming flights that can no longer utilize node A as an endpoint. Instead, those flights will have to be re-routed to airports near A. In our simulation, we will define a radius R such that all flights initially routed to A will instead be re-routed to another airport B in M, where M is the set of all airports within radius R of A. Any flight that is re-routed to B must then continue on its path through the network for the day from B.

Since flights will be re-routed, there will inevitably be flight delays that propagate through the network: planes will have to land at alternate airports when their destination is removed from the network; planes will be removed from the system when they are sitting inactive at an airport that shuts down, and replaced with existing planes at nearby airports that will have to take off at a later time. (This is in order to compensate for ground travel of expected passengers between airports and/or alter their predetermined paths to allow passengers from grounded planes to reach their intended destinations.)

2.2 Evaluation and Testing Methodology

The effectiveness of our rerouting algorithm will be determined by how dramatically delays propagate across the system, or more specifically, by a scalar value constructed from the sum of all flight delays across the system multiplied by the network centrality of the cut node to the network as a whole. The idea here is that more central nodes will likely cause larger propagations in delays across the system. For example, we cannot say that because cutting the airport in Duluth, Minnesota led to a smaller propagation of delay than cutting JFK International, the features we used to determine rerouting from Duluth were better. Scaling the sum of all delays against cut-node centrality should help resolve this issue.

We plan on determining which airports to re-route flights to (and which flights to re-route) by using a Support Vector Machine (SVM). We will parameterize key elements of the graph, such as the geographical distance from cut airport A to candidate airport B, the degree centrality of B, the betweenness centrality of B, etc. These features will serve in the training of our input vector, and the delay 'cost' will serve as the output value. After initially setting weights and running our simulation many times, the SVM should adjust the weights to provide the most optimal rerouting decisions given the chosen features.

2.3 The Data

Our initial data will consist of both real and generated data. The real data will be the actual daily flights that build our initial network. We plan on limiting this data to flights within the continental United States for simplicity. Our plan is to get a list of all flights in the continental United States by querying the Google Flights API. Instructions on how to do this can be found here: http://www.nohup.in/blog/using-json-google-flights.

Another website, OpenFlights.org, provides a massive database of all airports, airlines, and routes among airports around the world. We can use their dataset (available at http://openflights.org/data.html) to create a list of flights among airports in the continental United States. By querying each of those routes in the Google Flights API, we can then generate a list of flights and times on any given day in the United States. From this, we can then initialize the flight network for our simulation.

References

[1] Fleurquin, Pablo, Ramasco, Jose J. and Eguiluz, Victor M.; *Systemic delay propagation in the US airport network*; 2013.

[2] Fleurquin, Pablo, Ramasco, Jose J. and Eguiluz, Victor M.; *Characterization of delay propagation in the US air transportation network*; 2010.

[3] Pastor-Satorras, Romualdo and Vespignani, Alessandro; *Immunization of complex networks*; 2002.