

Data Visualization and Delay Prediction of Copa Airlines Flight Network

Sarthak Sharma (MS student), Dr Gordon Erlebacher (research adviser). Department of Scientific Computing, Florida State University.

Abstract: An event's delay is the difference between that event's scheduled and actual times. Delays in commercial flights can be in the aircraft's departure at the origin airport and arrival at the destination airport. These may cause considerable inconvenience, including passengers missing their next flight. This is especially true for airlines where many passengers transfer from one flight to another at the hub airport (e.g., Copa Airlines, which has PTY airport as its hub; "inbound" flights arrive at PTY, and "outbound" flights depart from PTY). Thus, delays can propagate in a flight network. Flight delays may be caused by many factors, some of which may be outside human control (e.g., bad weather). Hence, all delays can not be prevented. But if airlines and airports can predict flight delays with greater accuracy or certainty, then it may be possible to improve flight scheduling. Also, it may be possible to apply Artificial Intelligence to automate the flight scheduling process which is currently done manually in Copa Airlines. These developments may lead to better passenger experience and greater revenue. The first step for flight network analysis is to visualize the desired part of the flight network. We have developed a computational tool for data visualization of any desired part of the Copa Airlines flight network. Next, we are attempting to predict flight delays by applying Machine Learning techniques (e.g., artificial neural networks) to available flight data. In delay prediction, we can choose either classification (predicting delay category in 15-minute intervals) or regression (predicting the number of minutes in delay). The computational tools used include NumPy and Pandas for data processing; Plotly, NetworkX, and Matplotlib for data visualization; and TensorFlow and PyTorch for creating, training, and using artificial neural networks. All programming is in Python.



Figure 1: Copa Airlines flight hub network. Almost every flight of Copa Airlines has PTY airport as either origin or destination.

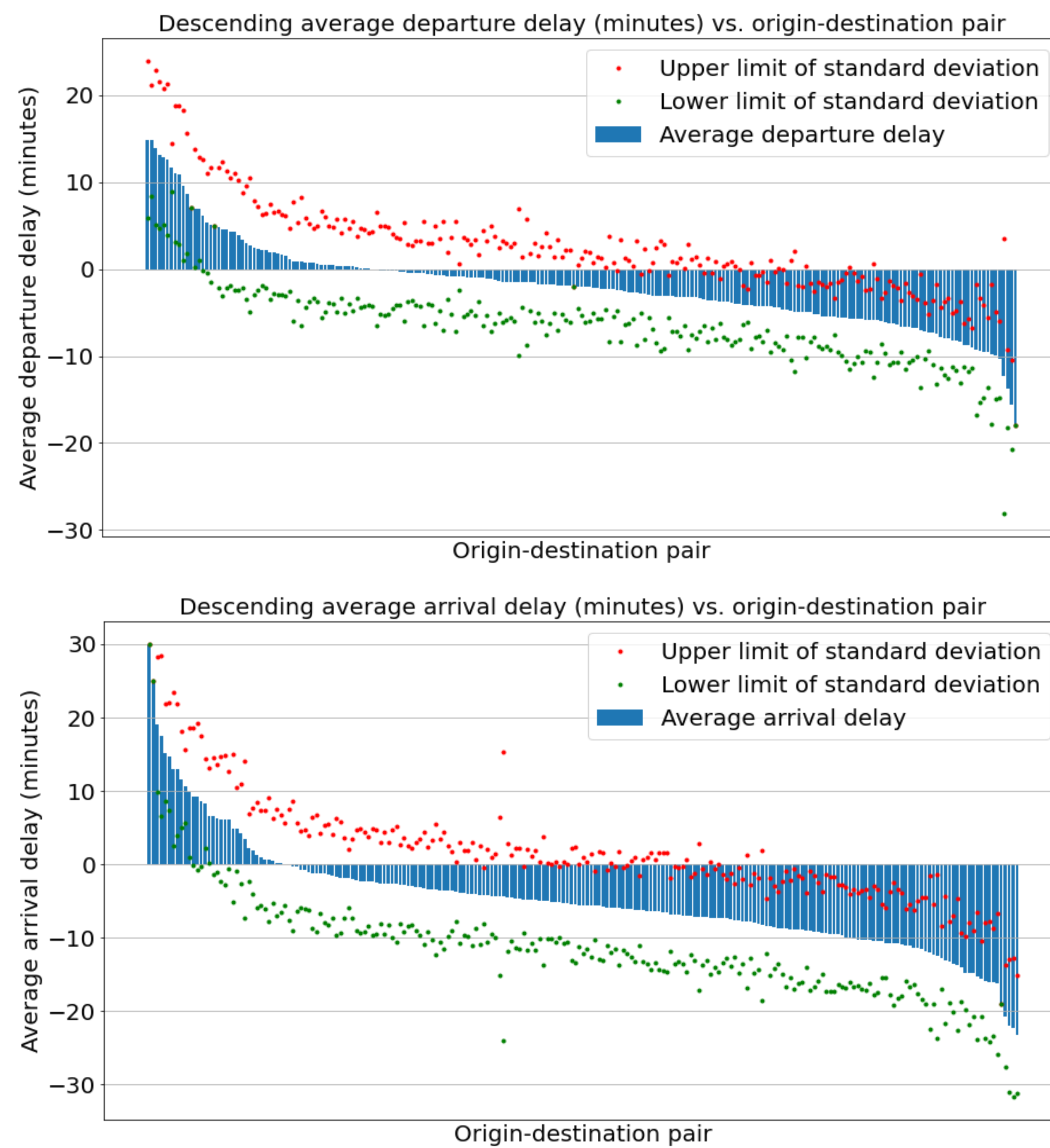


Figure 2: Average delays in departure and arrival arranged in descending order for various flight routes.

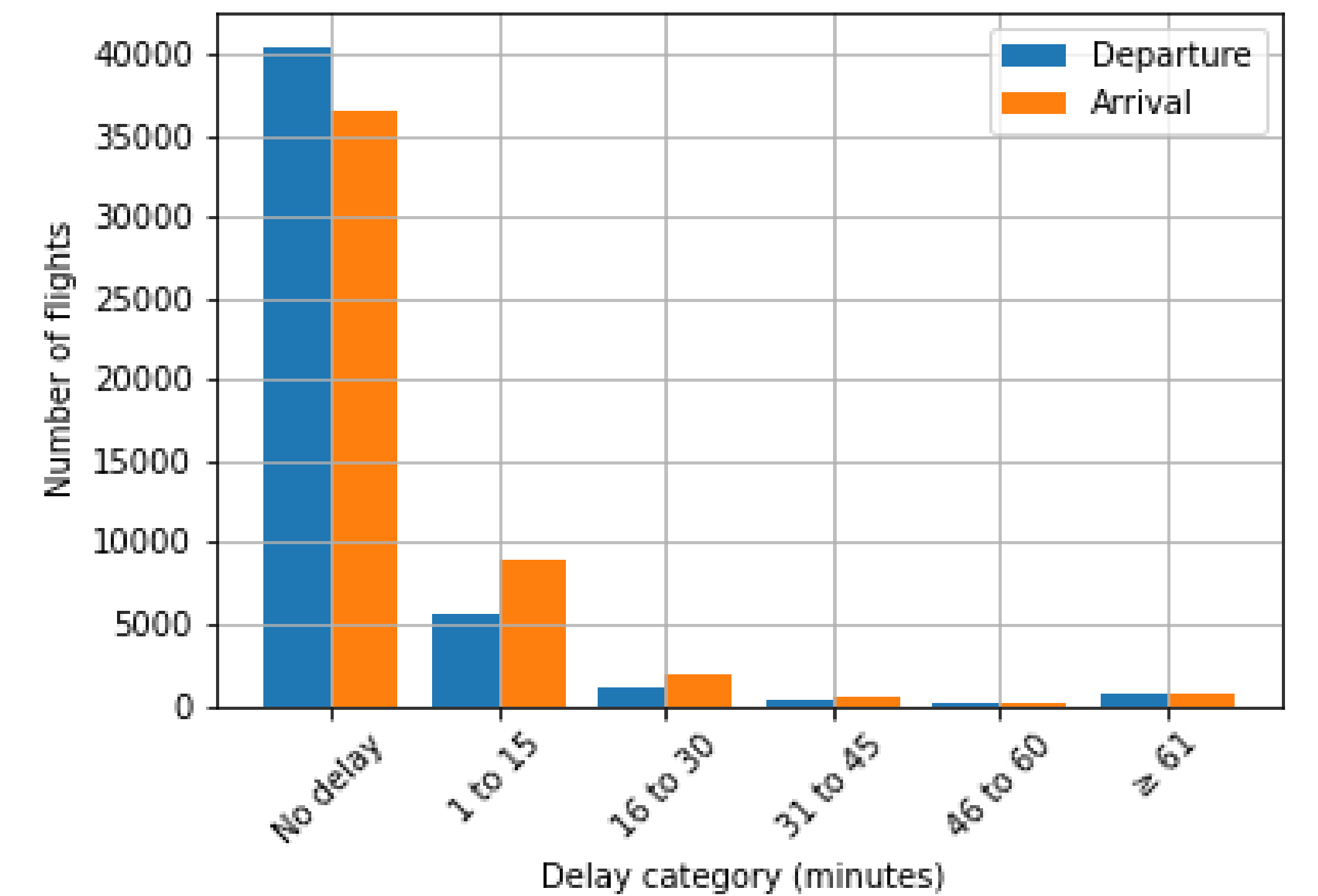


Figure 3: Copa Airlines average flight delays for departure and arrival, in different delay categories of 15-minute intervals. Most flights have no delays.

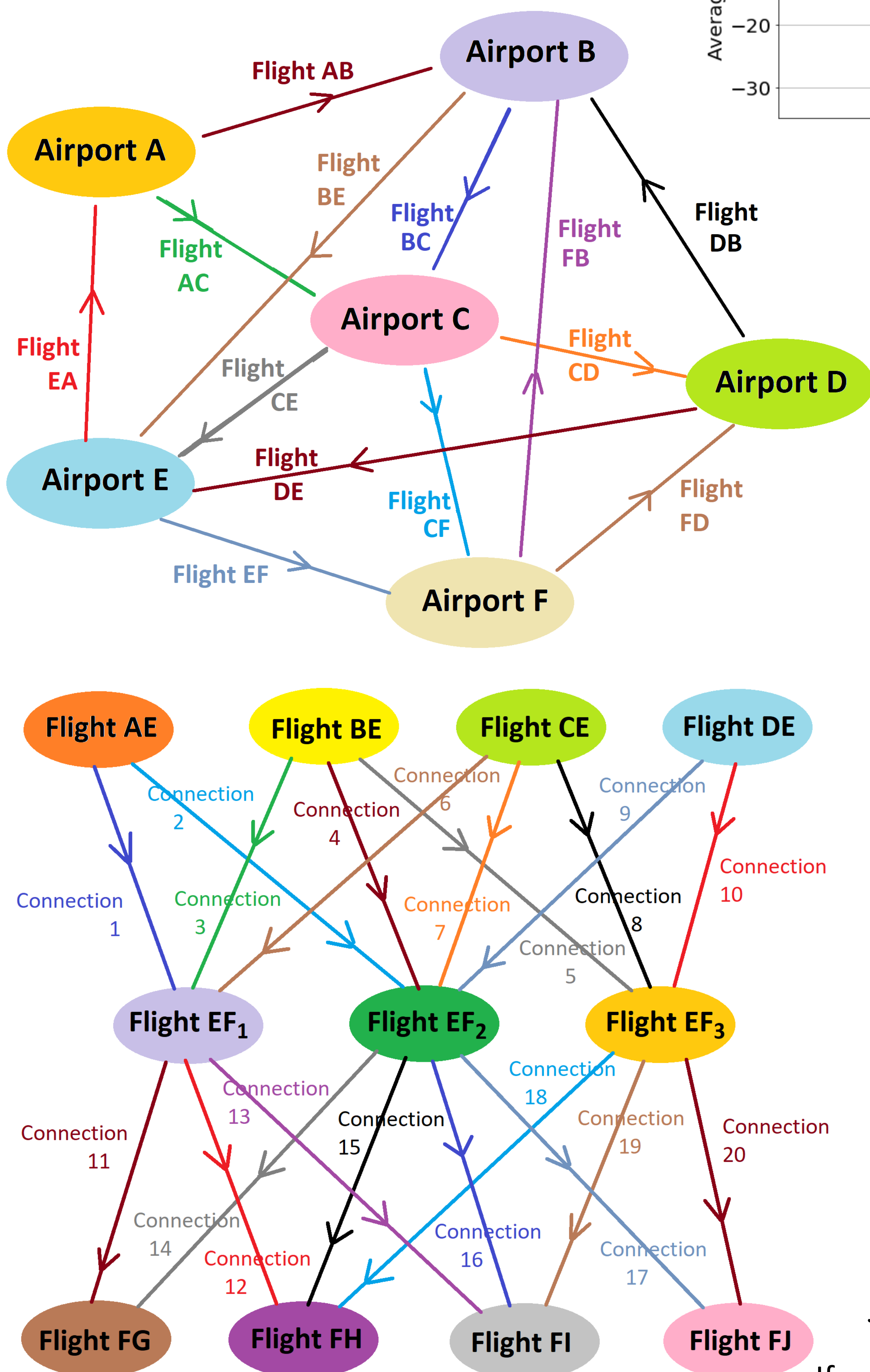


Figure 5: Two options for flight network visualization. Above: airports as nodes, and flights as links between nodes. Below: flights as nodes, and transfer of resources (passengers, crew, aircraft, etc.) as links between nodes.

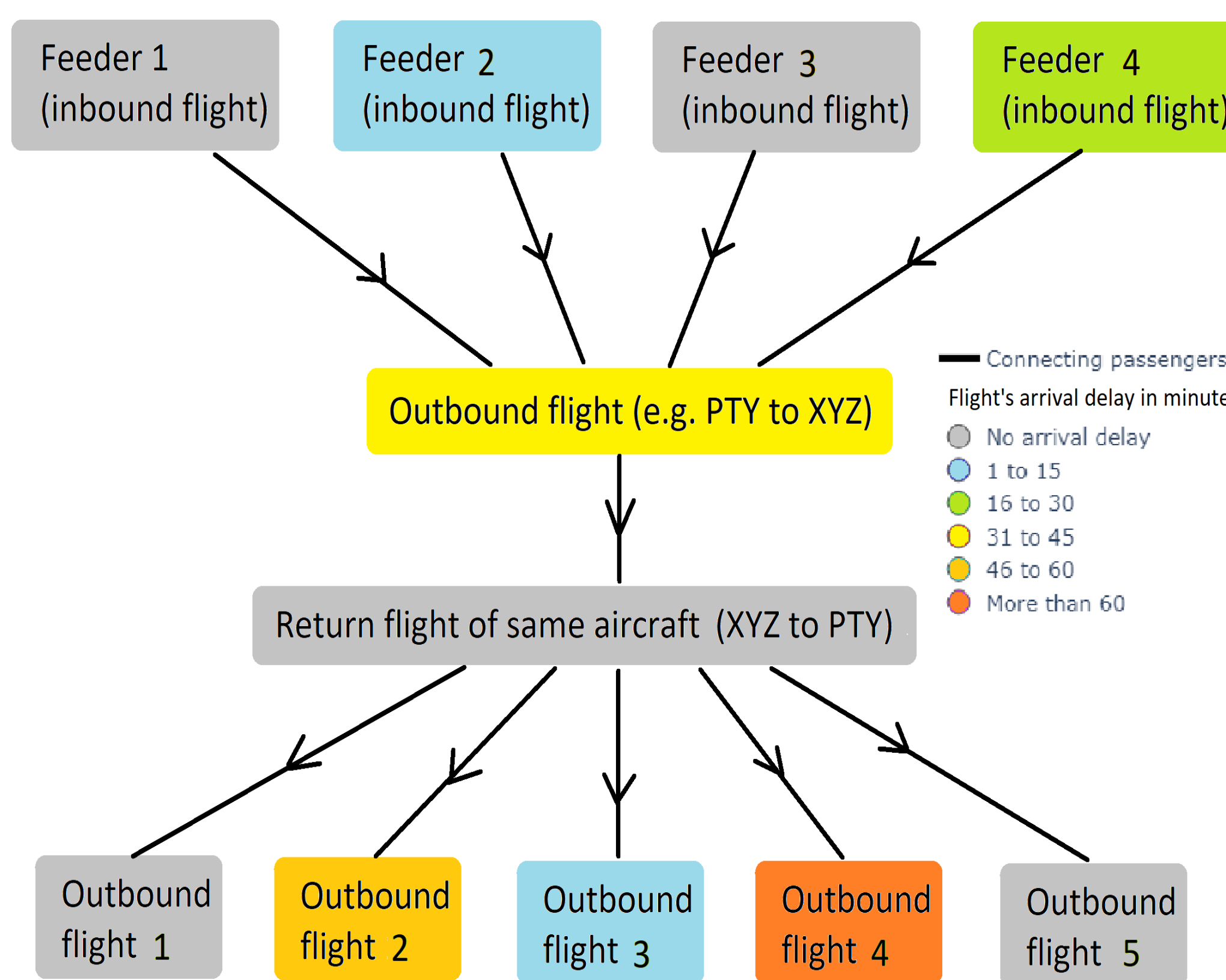


Figure 6: Flight network visualization scheme. Flights are depicted as nodes arranged in four levels. Each node is colored according to the respective flight's delay category. Resources (passengers, crew, aircraft, etc.) move from top to bottom. Selected outbound flight is at second level. First level (at the top) displays inbound flights feeding passengers to the selected outbound flight. Third level shows the return (inbound) flight of the same aircraft. Fourth level (at the bottom) shows the outbound flights that get their passengers from the return flight at the third level.

If a feeder (inbound) flight will arrive very late, then flight planners can either postpone departures of outbound flights getting their passengers from that delayed feeder flight, or put delayed connecting passengers on next available outbound flights to same destinations. Future work: flight delay prediction using graph neural network trained on past flight data.

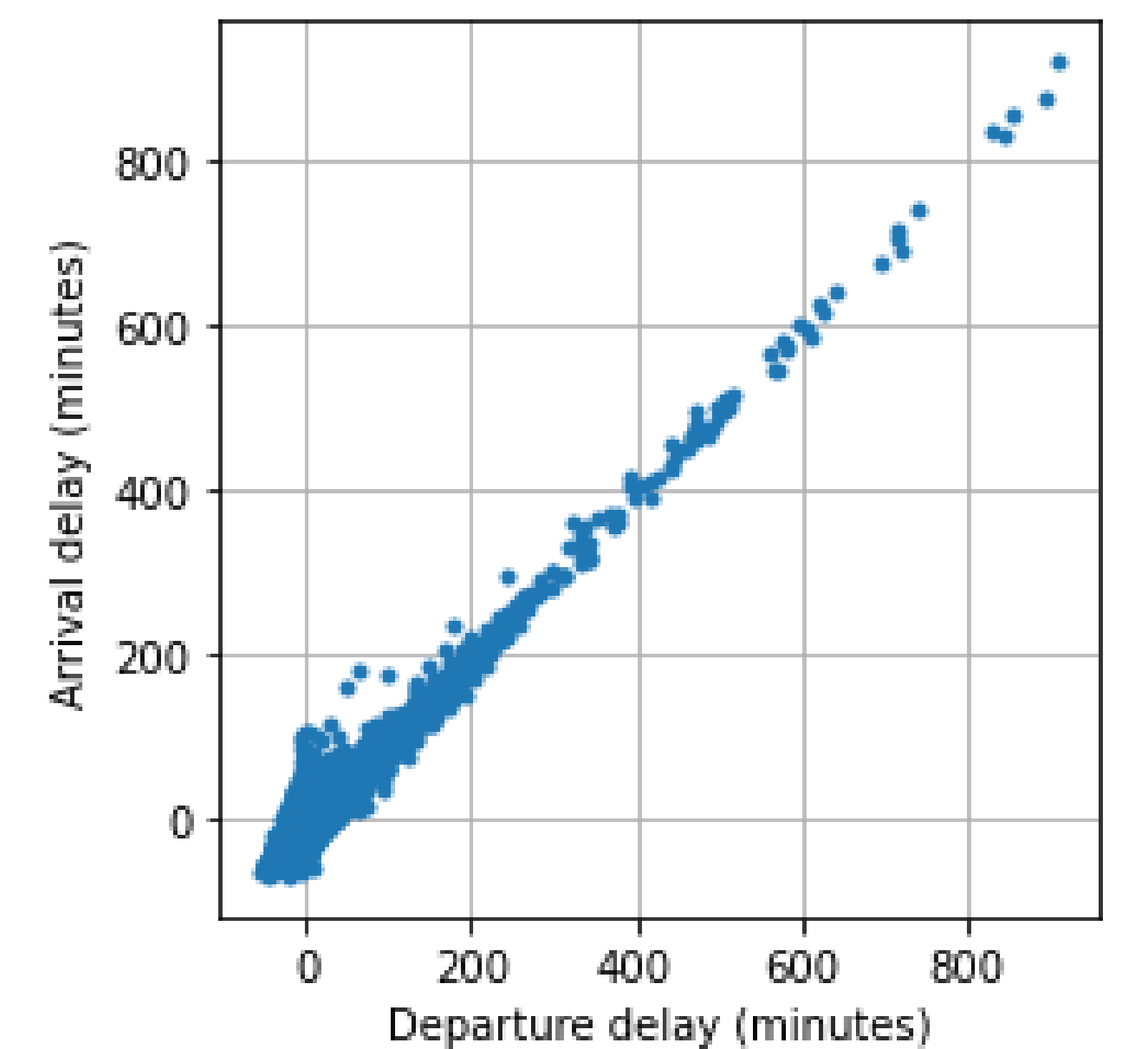
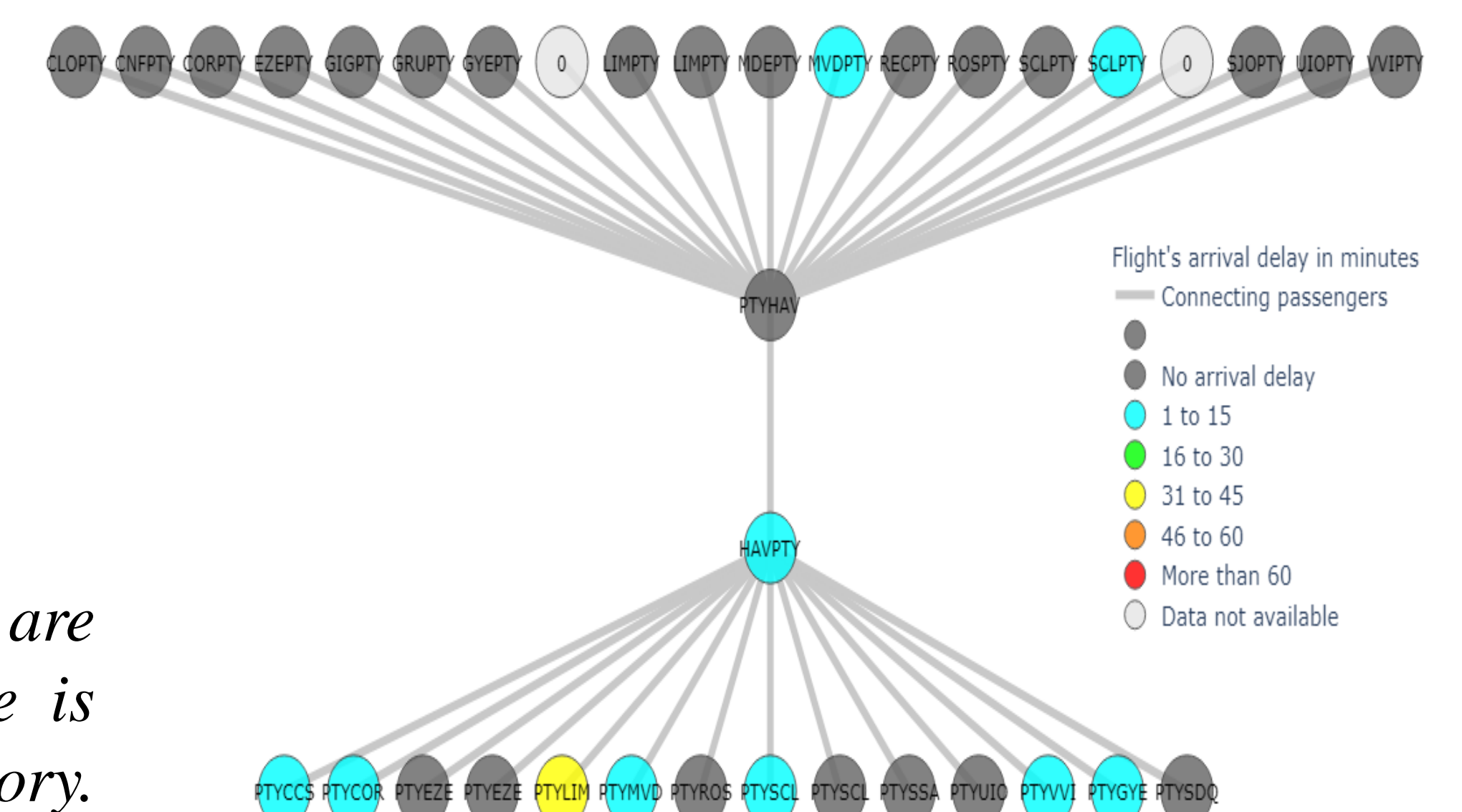


Figure 4: Departure delays and arrival delays are strongly correlated, with correlation coefficient of 0.9234. This is expected, as flights that depart late from origin, are likely to arrive late at destination.



Origin + destination code = PTYLIM
 Aircraft tail = HP1856
 Departure delay (minutes) = 18
 Arrival delay (minutes) = 32
 Flight number = 493
 Flight duration (hours) = 3.05
 PTY bank code = B4 Dep
 Scheduled departure date (Zulu) = 2019-10-01
 Scheduled departure time (Zulu) = 20:46
 Scheduled arrival date (Zulu) = 2019-10-02
 Scheduled arrival time (Zulu) = 00:22

Figure 7: Top: Example image generated by program, for selected flight (second level) from PTY to HAV. Bottom: flight data displayed as tool-tip (for node of flight from PTY to LIM, at bottom level).