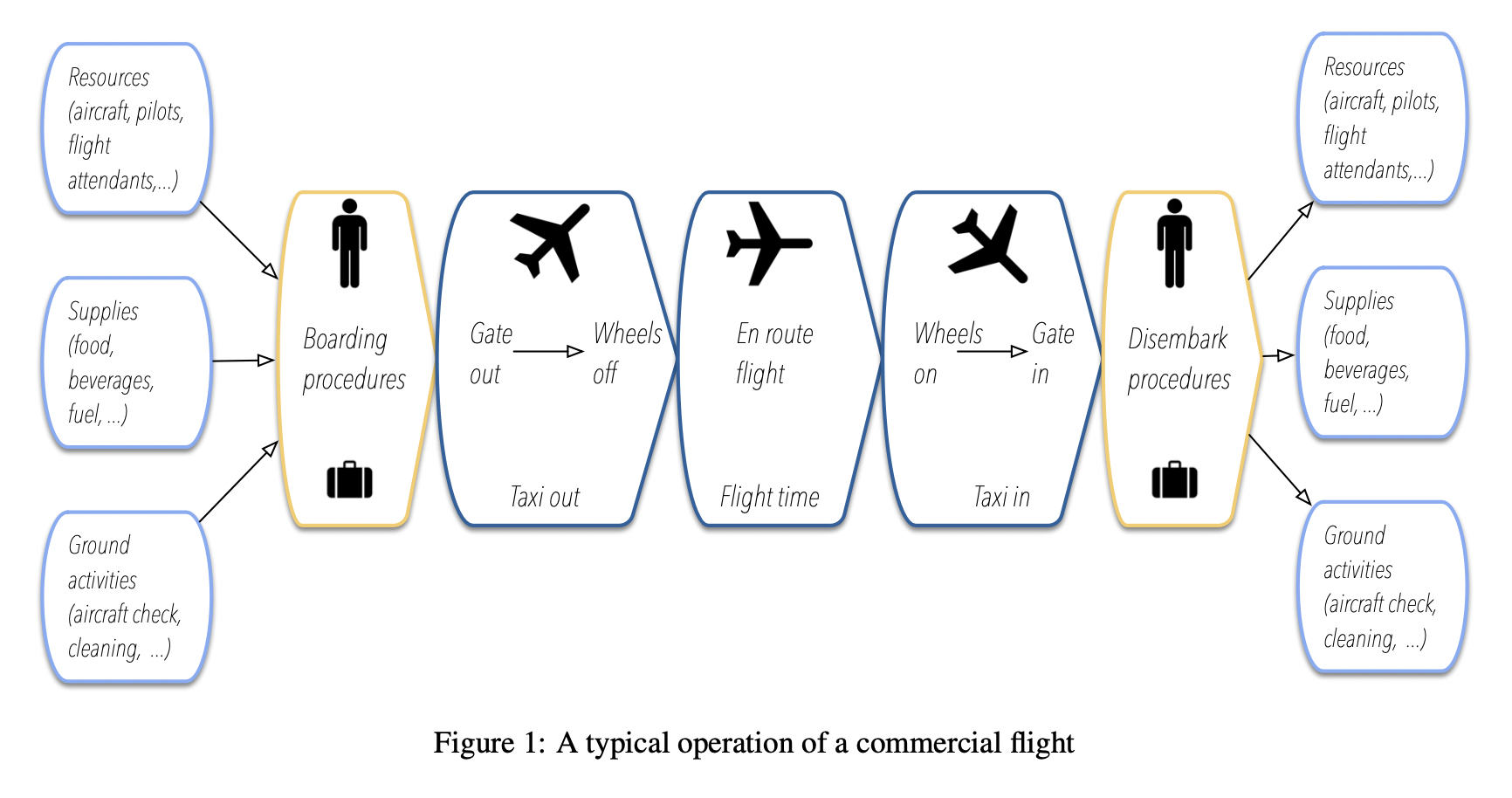
## A Review on Flight Delay Prediction

[carvalho\_review.pdf](https://drive.google.com/file/d/1r6Po1Oqh6X9UOUZXEcaYZuutCeLGFY4v/view?usp=sharing)

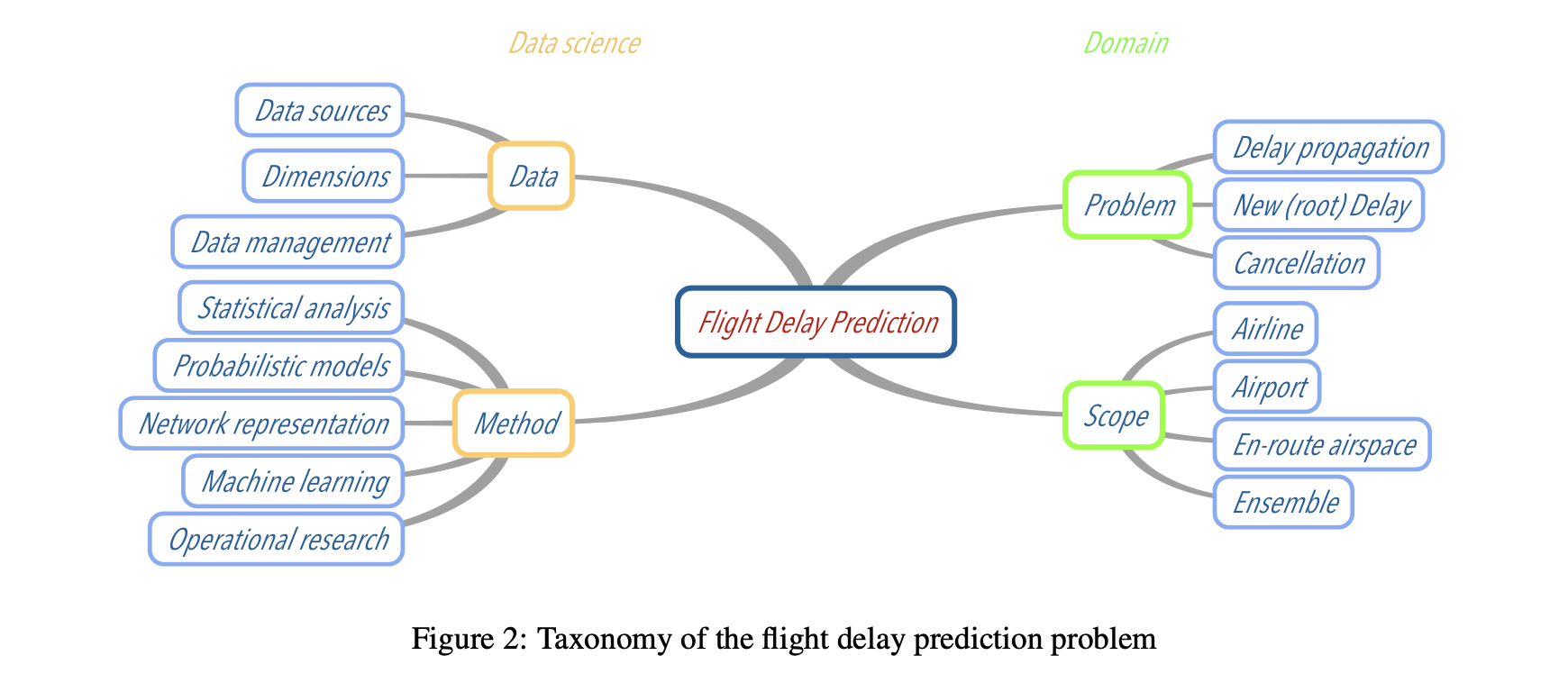
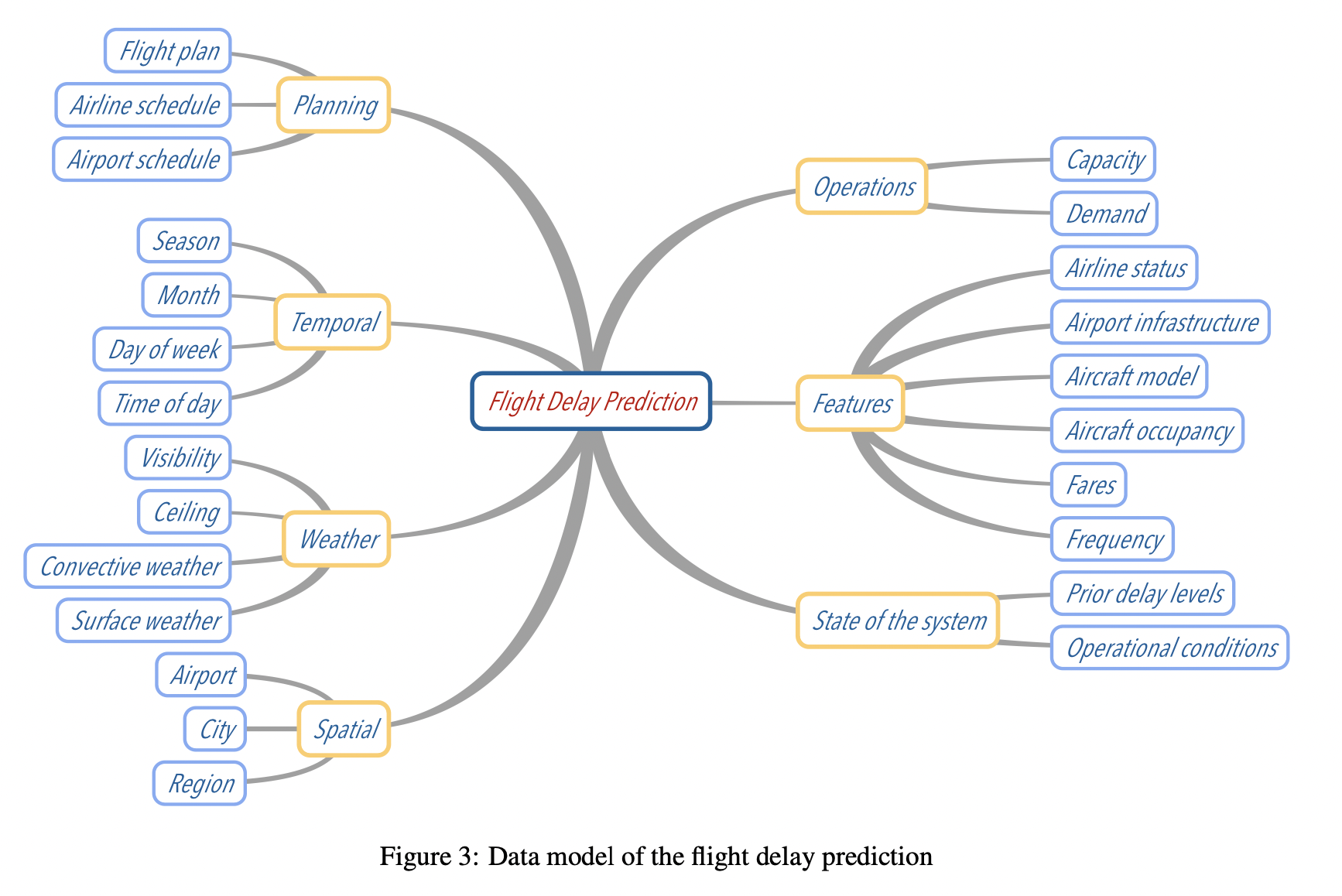
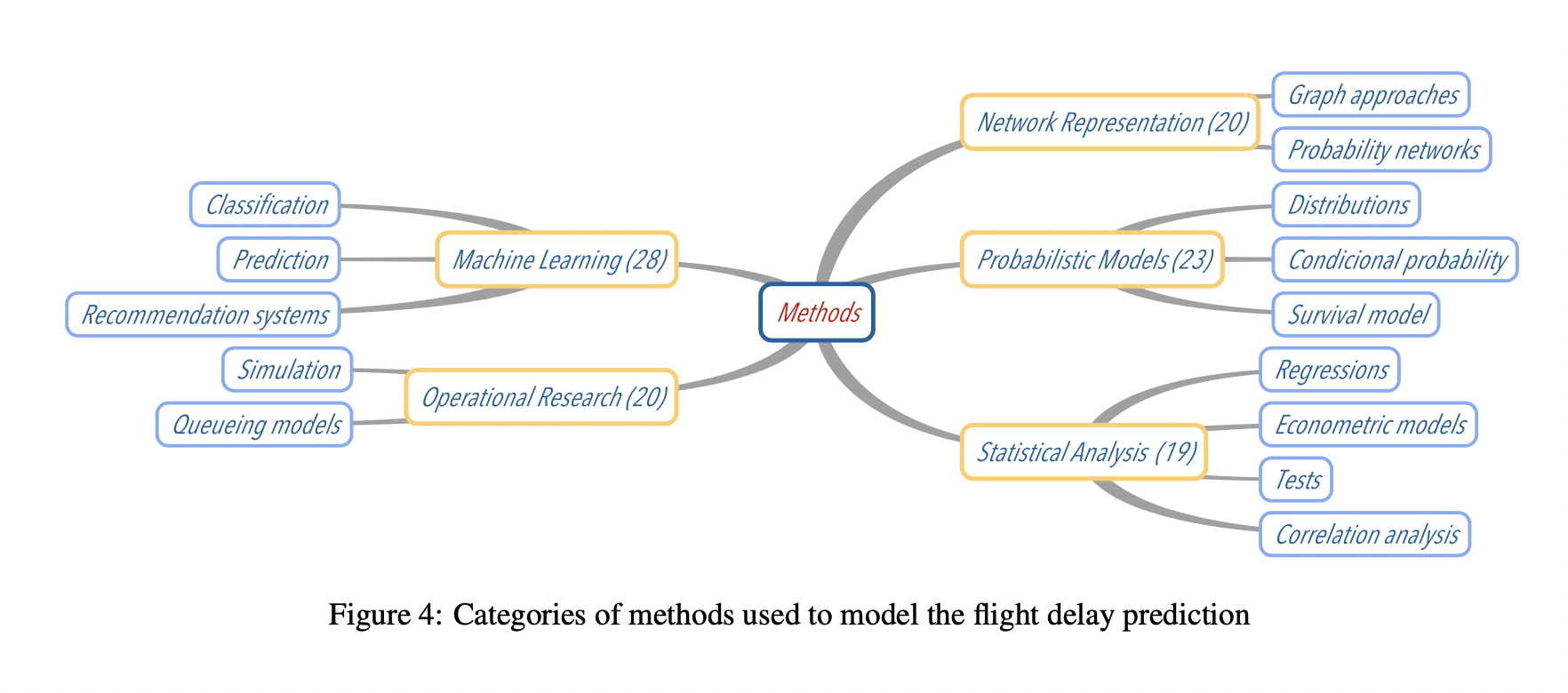
Introduction

* A delay may be represented by the difference between scheduled and real times of departure or arrival of a plane

The flight delay scenario

* Typical operation of a commercial flight
* Stages can take place at terminal boundaries, airports, runways, and airspace, being susceptible to different kinds of delays (ex. Mechanical problems, weather conditions, ground delays, air traffic control, runway queues, and capacity constraints)
* Any disruption in the system can impact the subsequent flights of the same airline
* Disturbances may cause congestion at airspace or other airports, creating queues and delaying some flights from other carriers
* 3 points of view of the flight delay prediction problem
  + Delay propagation: how delay propagates through the network of the transportation system
  + Root delay: predicting further delays and understanding their causes
  + Cancellation: delays can lead to cancellations, forcing airlines and passengers to reschedule their itineraries

Taxonomy

* Scopes, models, and ways of handling flight delay prediction problem
  + Flight domain features: problem, scope
  + Data science perspectives: data, methods 
* Problem: the core feature in domain taxonomy
  + Root delay and cancellation: create prediction models to tackle root delay, predicting when and where a delay will occur and what are its reasons and sources (ex. Models that seek to estimate the number of minutes, probability of level of delay for a specific flight, airline, or airport)
  + Delay Propagation: understand how delay propagates through airlines and airports based on the assumption that an initial delay has already occurred in the transportation system
    - Delays spread to other flights of the same airline as chain reactions
    - Important to measure how stable and reliable carriers can be to recover from delay propagation
* Scope
  + Delays can be induced by different sources and affect airports, airlines, en route airspace, or an ensemble of them
  + Assume a simplified system where only one of these actors or any combination of them is considered
  + Airports to predict delays → considered all airlines and en route airspace indifferently
    - Airports are also the focus when the objective is to investigate their efficiency based on delays of all carriers
* Data
  + Data sources
  + Dimensions → main input attributes for delay prediction models
  + Data management
    - One of the main tasks of data cleaning is outlier removal → extreme conditions may result in outliers that are not interesting if one is concerned about regular operations
    - Correlated and irrelevant attributes may provide model overfitting or decrease prediction performance
      * Better preprocessing = better prediction models
    - Data transformation: normalization, discretization, clustering
* Method
  + Statistical Analysis: use regression models, correlation analysis, econometric models
    - Delay multiplier and recursive models can help airlines to understand delay propagation effects through the network and to estimate costs of delays
  + Probabilistic models: estimate probability of an event based on historical data
    - Transition matrix to verify delay propagation to subsequent flights
    - Cancellation analysis → compute conditional probability to cancel a flight given that its previous flight was delayed
    - Normal distribution good for departure delays, Poisson distribution for en route/arrival delays
  + Network representation: study of flight systems according to a graph theory
    - Build DAGs to model the schedule of an airline
    - Use the shortest path algorithm to evaluation propagation effects
  + Operational Research
  + Machine Learning
    - KNN, neural networks, SVM, random forests for classification and prediction

Results and discussion

# A Comparative Analysis of Predicting Delays in Air Traffic Networks

[A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks.pdf](https://drive.google.com/file/d/1wPnv1PG5Tootj6KQ7UHcI20nGVzQO47c/view?usp=sharing)

Abstract

* Classification and Regression Trees
  + Prediction performance can vary significantly depending on the choice of model/algorithm and type of prediction
  + Need the right predictor variables (features)
* Markov Jump Linear System (MJLS) specifically designed to capture aggregate air traffic dynamics
* Tradeoff between model simplicity and prediction accuracy

Introduction

* In 2015, nearly 40% of delays were due to the delayed arrival of the incoming aircraft
  + High levels of interdependence in the delay dynamics
* Delays have been estimated to cost the US economy as much as $40 billion per year
* Artificial neural network (ANN) models are found to be effective for classification problems
* Background
  + Delays propagate in airspace systems due to multiple network interactions
  + Networked queuing models have been considered to understand the mechanism of delay propagation
  + Air traffic delay prediction

Problem Description

* Departure delay of a flight: difference between the actual time that an aircraft pushed back from the gate and its scheduled gate departure time
  + Assumed to be non-negative
* Origin-Destination (OD) pair delay

# Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms

[Prediction\_of\_weather-induced\_airline\_delays\_based\_on\_machine\_learning\_algorithms.pdf](https://drive.google.com/file/d/1_c2V1E9BOfCjncHD1Hj3ZST-YZnke8wJ/view?usp=sharing)