HAL open science

Impact of Covid-19 on passengers and airlines from passenger measurements: Managing customer satisfaction while putting the US Air Transportation System to sleep

Philippe Monmousseau, Aude Marzuoli, Eric Feron, Daniel Delahaye

▶ **To cite this version:**

Philippe Monmousseau, Aude Marzuoli, Eric Feron, Daniel Delahaye. Impact of Covid-19 on passengers and airlines from passenger measurements: Managing customer satisfaction while putting the US Air Transportation System to sleep. Transportation Research Interdisciplinary Perspectives, 2020, 7, pp.100179. 10.1016/j.trip.2020.100179 . hal-02935347

HAL Id: hal-02935347

https://enac.hal.science/hal-02935347v1

Submitted on 10 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Impact of Covid-19 on passengers and airlines from passenger measurements:

Managing customer satisfaction while putting the US Air Transportation System to sleep.

Philippe Monmousseau, Aude Marzuoli, Eric Feron and Daniel Delahaye

***Abstract*—The COVID-19 pandemic has had a significant impact on the air transportation system worldwide. This paper aims at analyzing the effect of the travel restriction measures implemented during the COVID-19 pandemic from a passenger perspective on the US air transportation system. Four metrics based on data generated by passengers and airlines on social media are proposed to measure how the travel restriction measures impacted the relation between passengers and airlines in close to real-time. The proposed metrics indicate that each airline has reacted differently to the COVID-19 travel restriction measures from a passenger perspective, therefore they can be used by airlines and passengers to improve their decision making process. This report comes ahead of official data related to the same sequence of events, thereby showing the value of passenger-borne data in an industry where corporate priorities, institutional prudence, and passenger satisfaction come close together.**

***Index Terms*—Air transportation system; passenger-generated data; passenger-centric metrics; COVID-19**

I. Motivation

A. The COVID-19 pandemic and the resulting travel restrictions from a US perspective

In response to the pandemic situation resulting from the outbreak of the corona disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), travel restrictions measures were implemented by various countries, impacting both domestic travel and international travel [1].

Italy was the first country to enforce a national lockdown [2] on March 9th 2020, after introducing on February 21st 2020 an initial measure confining only the northern region of Lodi. Two days after Italy's lockdown announcement, on March 11th 2020, the United States banned non-US travelers who had been to China, Iran and 26 member states of the European Union (EU) to enter the US, and later extended the ban to non-US travelers who had visited the United Kingdom and Ireland on March 16th 2020 [1]. The EU officially closed the external borders of 26 of its member states to nearly all non-EU residents on March 17th 2020 [1]. On March 19th 2020, the US Department of State issued a Level 4 Global Health Travel Advisory, which cautions all US citizens against international travel, still in place as of May 6th 2020 [3].

This dramatic sequence of events forms the thread against which the air transportation system has had to progressively put itself to a semi-comatose state to address fast-growing sanitary and economic concerns. For these reasons, the following dates are indicated with dotted lines in every graph throughout this paper in order to better visualize the timeline of each figure.

1) The **Lodi** region lockdown in Italy: February 21st, 2020
2) **Italy**'s lockdown: March 9th, 2020
3) **US ban** of non-US travelers from the EU, China and Iran: March 11th, 2020
4) **EU** external border **closure**: March 17th, 2020
5) **US Level 4** Global Health Travel Advisory: March 19th, 2020

Figure 1 presents the number of passengers arriving at US immigration across all airports of entry using the "Airport Wait Times" data from the Customs and Border Protection (CBP) website [4]. This plot illustrates clearly the effect of these travel restriction measures on the international traffic coming to the US. For a more detailed presentation of the available CBP dataset, the authors recommend reading [5], which also presents an analysis of the wait times at US airport immigration services from January 2013 to January 2019.

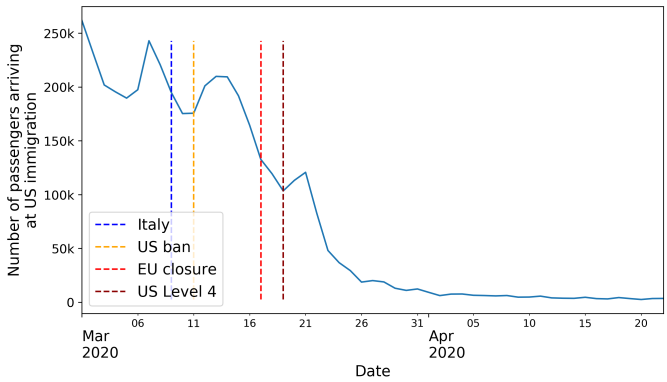


Figure 1: Evolution of the daily number of passengers arriving at all US airports of entry from CBP data.

The air transportation system is an essential system to understand and to study under pandemic situations from various perspectives, e.g. the propagation of diseases inside airplanes [6], the propagation of epidemics via airplanes [7], or the

effect of the travel restrictions on airline employment [8]. This paper focuses on the effect of the pandemic on the attitude of passengers towards airlines. In 2019, considering eight major US airlines and thirty-four major US airports, Twitter users wrote a median of 13,255 tweets mentioning an airport and a median of 295,904 tweets mentioning an airline, indicating that users interact more with airlines than with airports.

B. The limitations of traditional approaches to assess the impact of COVID-19 on the air transportation system

The travel restrictions, and the other measures taken by a majority of countries worldwide, are having an unprecedented impact on the air transportation system. Until official flight data are released in the United States regarding international and domestic air transportation there are no means of measuring this impact on the US air transportation system, except by relying on non-traditional data sources.

Traditionally, the metrics used to measure the state of the US air transportation system are focused on flight performances, such as the amount of delay per flight, the number of delayed flights, the number of cancelled flights and the number of carried passengers. The data considered for these metrics are gathered by the US Department of Transportation Bureau of Transportation Statistics (BTS) [9]. The data are first processed by airlines and airports and then provided to the BTS, which then publishes the data as a monthly report. The BTS reports pertaining to on-time flight data are usually published with a latency of two months. This latency is not well adapted for monitoring and analyzing the effects of situations such as the COVID-19 pandemic on the US air transportation system.

Figure 2 presents the number of international and domestic flights from March 1st 2020 to April 22nd 2020 using available data from CBP and from BTS as of June 24th 2020. From these data, the number of daily domestic flights drops by half in the second half of March 2020 but no conclusion can be drawn for the month of April 2020. While not technically asleep,

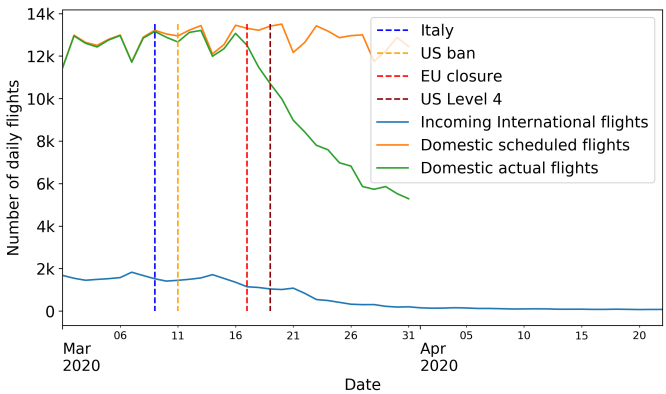


Figure 2: Evolution of the daily number of arriving international flights from CBP data, as well as the daily number of scheduled and actual US domestic flights from BTS data.

many flights kept being flown by airlines because they feared to lose their slots [10], [11] or because they had to keep flying routes in order to receive financial aid [12], a situation close to a sleep condition called ”nightmare”.

Additionally, various studies have shown that passengers were disproportionally impacted by flight capacity reduction [13], [14], [15], [16], highlighting the differences between measuring flight delays and flight cancellations and measuring the actual passenger delay. For example, based on data from a major US airline, they show that disrupted passengers, whose journey was interrupted by a capacity reduction, are only 3% of the total passengers, but suffer 39% of the total passenger delay. The necessity of adding a passenger-centric approach when evaluating the air transportation system was later put forward by NextGen [17] in the US and by ACARE Flight-Path 2050 in Europe [18]. A first attempt at implementing passenger-oriented metrics was performed by Cook et al. [19]. Integrating passenger objectives in airport decision making processes was introduced within the concept of Multimodal, Efficient Transportation in Airports and Collaborative Decision Making (META-CDM) [20], [21], [22]. Though these works give an important place to passengers, they still heavily rely on flight-centric data and have thus the same latency limitation.

Several years later, the advocated shift from flight-centric metrics to passenger-centric metrics still has to be actually implemented by the governing agencies. In a report published in 2016, EUROCONTROL and the FAA presented metrics regarding punctuality that combines airline and passenger views into a single view [23].

Already in 1992, [24] advocated for the need of unified airport performance measures that would balance the expectations of passengers, airlines and airports along with the expectations of other actors (such as restaurants or governments). Understanding the passenger experience, or at least the passenger perception of airport and airline quality has since been the focus of many studies. [25] first proposed to introduce surveys based on fuzzy set theory in order to analyze airline service quality. [26] performed a thorough survey of airline perception related studies from 1995 to 2006, pointing out the decrease in customer service throughout the airline industries. For more informations on the various survey-based methods used, [27] conducted a survey of survey based analysis of public transportation system. They concluded that even though researchers keep trying to improve the complexity of the models to better model passenger satisfaction of a public transportation system, managers and practitioners use simpler models in order to reach their goal of improving passenger perceived service quality for an increase of income.

Passenger surveys conducted at airports for airports or airlines, while very detailed, remain limited to very small samples of passengers and short time periods, and may not be representative. For example, among some of the founding survey studies, [25] have a sample size of 211 passengers and [28] have a sample size of 385 passengers. They are also expensive and time consuming to implement, making their use for measuring the effects of major perturbations, such as the COVID-19 pandemic, on the air transportation system

cumbersome and difficult to update.

C. Passengers as sensors of the air transportation system

Using passenger-generated data in order to analyze the efficiency of the air transportation system was made easier thanks to the ubiquity of smartphones. Data from WiFi hotspots and Bluetooth beacons, along with historical data, are used to analyze passenger behavior at airports [29], [30] and at transit stations [31]. If available, data generated by passengers smartphone and collected by phone carriers can be processed to analyze the door-to-door behavior of passengers [32], [33], [34], both under nominal and degraded conditions. However data gathered directly from smartphones are proprietary data and are not often publicly available for research.

This paper proposes an alternative approach to analyzing the air transportation system by focusing on airline performances with respect to their passengers using data generated by airlines and by passengers. The importance for airlines of improving the waiting environment at airports in order to improve passenger satisfaction is already highlighted in [35] and is generalized for riders at transit stations in [36]. In the specific case of US air transportation, Twitter is an important medium for direct communication between passengers and airlines. For example, over the month of January 2020, more than 300 tweets were written on average every day by the customer services of four major US carriers (Southwest Airlines, Delta Airlines, American Airlines and United Airlines) and more than 800 tweets were written on average every day by their customers. This direct communication is a form of unsolicited feedback from customers and is therefore inherently biased, towards both extreme dissatisfaction and extreme satisfaction. However, [37] suggests that continuous quality monitoring can benefit from the extreme responses contained in unsolicited feedback. For example, unsolicited feedback within social media activity is used by [38] to detect information about defective components in the automotive industry. In Europe, KLM promised a 30-minute customer-response time in the afterwake of the air transportation major disruption initiated by the eruption of an Icelandic volcano in 2010 [39].

The real-time availability of Twitter data is the starting point of many studies of large scale events, such as natural disasters, and how Twitter could be used to help emergency responders [40], [41], [42], [43], [44], [45]. Regarding applications to the air transportation field, most works mining Twitter data focus on creating and improving airline sentiment classification methods [46], [47], without proposing any direct use of their results to improve airline service or passenger satisfaction.

Applications of sentiment analysis for airlines are proposed by [48] who use sentiment and topic analysis to extract from around a thousand tweets the information needed to calculate a proxy of the Airline Quality Rating, a flight centric metric including a measure of customer complaints introduced by [49]. [50] analyze airline customer service experiences both by manually labelling tweets related to four airlines written on five different days of 2010 and containing one of three keywords (”good”, ”fail” and ”lounge”) into six categories (personal, positive, negative, promotion, question or news). Tweets within the positive and negative categories are then analyzed to determine which airline services are associated with positive or negative sentiments. [51] show that airlines are more likely to respond to customers with greater popularity, and have a tendency to respond more to complaints than to compliments, where complaints and compliments are determined based on a set of manually defined keywords. These studies show that Twitter can be used by airlines in order to gain some insights on how passengers perceive their service and how they treat their passengers on Twitter.

Using Twitter to build a real-time estimator of the air transportation system is investigated in [52], [**?**] whose purpose is to estimate flight-centric values per airport before they were released by BTS. This paper takes another approach and proposes several passenger-centric metrics constructed from passenger-generated data in order to offer a passenger-centric perspective of the air transportation system, with a focus on the relation between airlines and passengers. This paper is not directly interested in ”classical” measures of performance, such as those of direct interest to airlines (productivity, profitability) or to Air Navigation Service Providers (on-time performance and other metrics used to evaluate technical development programs, such as NextGen in the US or SESAR in Europe). The proposed work introduces measures of satisfaction and feedback expressed by the passengers themselves. Such measures are complementary with and different from the foregoing, although correlations may exist.

Eight airlines, and their associated Twitter handles, are considered in the analysis below: American Airlines (@AmericanAir), Delta Air Lines (@Delta), United Airlines (@united), Alaska Airlines (@AlaskaAir), Southwest Airlines (@SouthwestAir), JetBlue Airways (@JetBlue), Spirit Airlines (@SpiritAirlines) and Frontier Airlines (@FlyFrontier and @FrontierCare). The first four are legacy airlines, and the last four are low-cost carriers. All tweets written from these airlines Twitter accounts were scraped from February 16th 2020 to May 3rd 2020 and are categorized as ”customer service tweets”. This category contains both public replies to customers and public tweets for everyone to read. All tweets written over that same period and mentioning at least one of the airline handles that was not written from the corresponding airline Twitter account were also scraped and categorized as ”passenger tweets”.

The rest of this paper is structured as follows: Section II describes the first two metrics based on a Twitter sentiment analysis and how they can be used in light of the COVID-19 situation. Section III then describes two additional metrics based on selected keywords and how they can be used to assess the performance of airline communication during the COVID-19 pandemic. Section IV concludes this paper and discusses future research directions.

II. Impact of the COVID-19 on airline and passenger mood

A. Evaluating the mood expressed in tweets

A first step in sentiment analysis is to create a labelled dataset containing an equal number of tweets expressing a *positive* sentiment and tweets expressing a *negative* sentiment. The training dataset used in this study is based on the works of [53], [54]. Emoji filters are used to extract 49,030 tweets written in 2017 by airlines and their customers and automatically assign a positive or negative sentiment label to each tweet according to Table I.

Table I: Emoji sentiment association

Category	Emojis
Positive	":)", "=)", ":-)", ";)", ";-)", ":-D", ":D", "=D"
Negative	":(", ":-(", "=(", ":-@", ":'(", ":-|"

A processing pipeline is then applied to each tweet in order to transform the text contained within each tweet into a vector of tokens that can be processed by the sentiment classifiers. A token is either a generic keyword, a single word, a bigram or a trigram. Bigrams (resp. trigrams) are combinations of two (resp. three) consecutive words that are commonly used together, e.g. "do not like" is a trigram. In order to reduce the sparsity of the considered vocabulary, generic keywords are used to replace mentions to other Twitter users ("@someone" becomes "MENTION") and mentions to the considered airlines (e.g. "@united" becomes "AIRLINE"). They are also used to replace date related association of words, e.g. "March 11th 2020" becomes "DATE" and "8am" becomes "TIME". Generic keywords are also used to indicate if the tweet contains a link to a website or if a picture is embedded in the tweet. To remove any potential bias of emojis on the learning process, since every tweet in the training dataset contains an emoji, all emojis are replaced by the keyword "EMOJI".

Since the text contained in a tweet can be loosely written, with emphasis given to words with repeated letters such as "looooove", the number of duplicate letters were limited to two in every word: "looooove" becomes "loove". Negative bigrams are also automatically created by merging negation words ("no", "not" and "never") with the word that follows it. Once all the tokens were created, the tokens occurring in fewer than twenty tweets within the training dataset are removed as well as the tokens appearing in more than 75% of the tweets within the training dataset.

Five classifiers are trained on the training dataset and then tested on the labeled dataset provided by [55]: an AdaBoost classifier [56], a gradient boosting classifier [57], a random forest classifier [58], a naive Bayesian classifier [59] and a logistic regressor [60] using the scikit-learn python library [61].

Each classifier gives a score of 1 if it considers that the tweet expresses a positive sentiment and a score of 0 if it expresses a negative sentiment. In effect, each classifier calculates a predicted probability for a tweet of being positive, and then rounds that predicted probability to the closest integer (0 or 1). The classifiers are transformed into regressors by considering the probability for a tweet of being classified as positive. The output of all trained regressors is then averaged into one single score ranging from 0 to 1, with a score of 0 indicating a negative mood and a score of 1 indicating a positive mood.

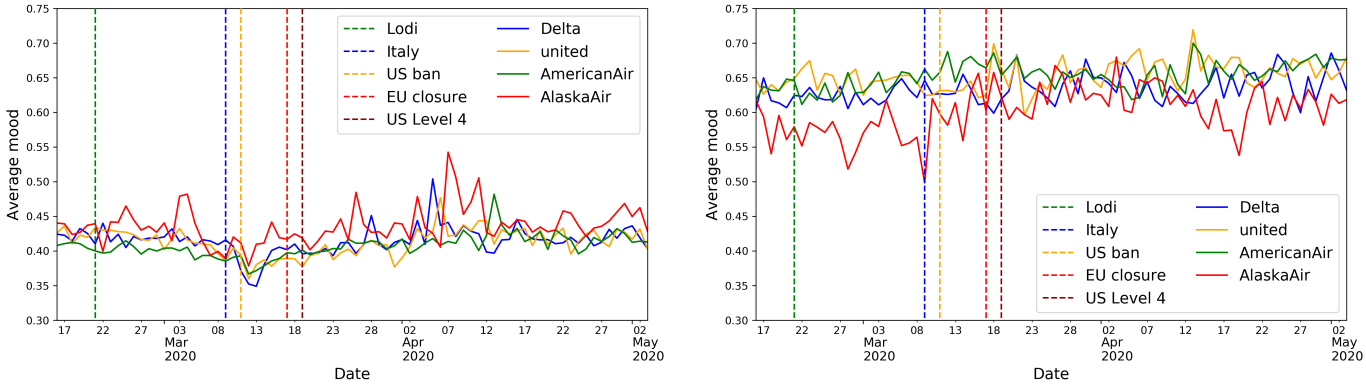
B. Daily mood evolution

Once the sentiment expressed within each tweet is averaged on a daily level, the effect of the travel restriction measures on the expressed passenger mood can be compared with their effects on the expressed airline mood. Legacy airlines are usually considered as offering a higher quality service to customers than low-cost carriers, with an average of close to 296 tweets written a day by the customer service of the four considered US legacy airlines versus an average of 112 tweets written a day by the customer service a day for the four considered low-cost carriers. The evolution of the mood expressed by passengers and airline customer services is presented in the following subsections, first for the legacy airlines and then for the low-cost carriers.

1) Case of legacy airlines: Figure 3 shows the evolution of the mood expressed by the four legacy airlines considered and by their passengers from February 16th 2020 to May 3rd 2020.

From Figure 3(a), a drop in the mood expressed by passengers can be observed starting right after the Lodi lockdown with a steep decrease right after the US travel ban for the three major airlines (Delta Air Lines, United Airlines and American Airlines). The sentiment extracted from the tweets from Delta's passengers has the steepest descent but also the sharpest recovery. The case of Alaska Airlines exhibits special characteristics: a #AlaskaHappyHour campaign, giving Twitter users the opportunity of winning a free flight to Alaska, was taking place early March 2020. This campaign could explain why the expressed mood in passenger tweets increased between March 1st 2020 and March 5th 2020 and could have compensated a potential decrease in the passenger expressed mood linked to the travel ban announcement.

Regarding the mood expressed in tweets written by the airline customer services, shown in Figure 3(b), it only decreases for Delta Air Lines and United Airlines starting at the announcement of Italy's lockdown. An opposite reaction is seen with the mood expressed by American Airlines customer service, which increases over that same period. Comparing Figure 3(a) and Figure 3(b) shows that Delta Air Lines and Alaska Airlines have the highest expressed mood on average within their passenger tweets over the considered period, but the lowest expressed mood within their customer service tweets of the four legacy airlines. An explanation of the better mood expressed by their passengers could be that these airlines expressed a mood closer to their passengers' actual mood. A gap between the mood extracted from passenger tweets and the mood extracted from airline customer service tweets is visible from one figure to another, with airline customer



(a) From passengers of major airlines

(b) From customer service

Figure 3: Daily average mood expressed in tweets containing airline Twitter handles for four legacy airlines between February 16th 2020 and May 3rd 2020.

service tweets expressing a mood about 0.2 points higher than passenger tweets.

2) Case of low-cost carriers: Similar conclusions can be drawn when analyzing the mood associated to tweets from passengers and customer services of low-cost carriers. Figure 4 shows the evolution of the expressed mood from February 16th 2020 and May 3rd 2020 in the passenger and customer service tweets of the three low-cost carriers considered.

Figure 4(a) indicates that the mood expressed by Spirit Airlines passengers and by Frontier Airlines passengers is significantly lower on average than the mood expressed by passengers of JetBlue Airways and Southwest Airlines over the months of February and March 2020. There is a spike in the mood extracted from tweets written by JetBlue passengers around March 26th 2020. This date is also the day when the governor of New York thanked JetBlue for offering free flights to health care workers in order to help the state handle the spread of COVID-19[1]. It also corresponds to the period when an update of their mobile application contained the message "Now, go wash your hands", prompting an amused reaction of their passengers. The drop in the mood expressed in the tweets written by legacy airline passengers after Italy's lockdown is less visible in the tweets written by passengers of low-cost carriers, with the exception of the mood expressed by passengers of Southwest Airlines.

Looking at the mood expressed by low-cost carrier customer services presented in Figure 4(a), the mood expressed by the customer service of Frontier Airlines displays a highly varying behavior, oscillating between 0.23 and 0.83 with discontinuities since on certain days no tweets were written by their customer service. For the other three low-cost carriers, the gap between the mood extracted from the tweets written by Southwest Airlines customer service and the mood extracted from the tweets written by the customer services of the other two carriers reduces significantly the day after Italy's lockdown. Similarly as for legacy airlines, a gap of about 0.2 points is visible between the mood expressed within passenger tweets and airline customer service tweets by comparing Figure 4(a) and Figure 4(b).

[1]https://twitter.com/NYGovCuomo/status/1242941085535608835

C. Passenger-centric metrics

Based on the observations presented in Section II-B, two passenger-centric metrics are proposed to measure the relation between airline customer services and their passengers. The first proposed metric aims at measuring the evolution of the airline mood relative to the mood of their passengers. Diverging mood evolutions are given a low score: if the average mood expressed by passengers is decreasing, the average mood expressed in the tweets written by the airline customer service should not be increasing.

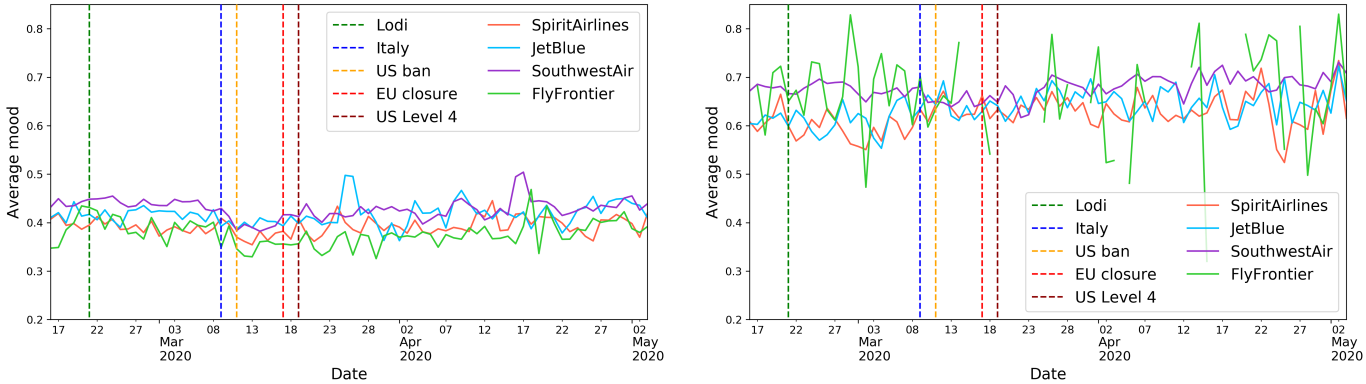
Proposed passenger-centric metric 1: The airline **empathy score** is defined as the Pearson correlation between the evolution of the average mood expressed by passengers in their tweets and the evolution of the average mood expressed by the airline customer service in their tweets.

The empathy score Ξ is calculated using the following formula:

$$\Xi = \frac{\sum_i (p_i - \bar{p})(c_i - \bar{c})}{\sqrt{\sum_i (p_i - \bar{p})^2 \sum_i (c_i - \bar{c})^2}} \quad (1)$$

where the set $\{p_i\}_i$ (resp. $\{c_i\}_i$) is the ordered set of the daily expressed mood in passenger tweets (resp. in airline customer service tweets), and $\bar{p}$ (resp. $\bar{c}$) is the average daily expressed mood over the considered period in passenger tweets (resp. in airline customer service tweets).

The empathy score Ξ goes from -1 to 1, with a score of 1 meaning that the airline customer service expressed mood is in agreement with the mood expressed by their passengers. On the opposite, a score of -1 indicates that the mood expressed by the airline customer service is in complete opposition of phase with the mood expressed by their passengers. Such a score would indicate that the mood expressed by the airline customer service increases when the mood expressed in passenger tweets decreases, and *vice-versa*. A score of 0 indicates that



(a) From passengers

(b) From customer service

Figure 4: Daily average mood expressed in tweets containing airline Twitter handles for three low-cost airlines between February 16th 2020 and May 3rd 2020.

the mood expressed by the airline customer service and the mood expressed by their passengers are uncorrelated.

The second proposed metric aims at measuring the gap observed between the mood expressed by passengers in their tweets and the mood expressed in the tweets written by airline customer services.

Proposed passenger-centric metric 2: The airline **sentiment gap** is the average difference between the mood expressed by passengers and the mood expressed by airlines.

The airline sentiment gap Δ is calculated using the following formula:

$$\Delta = \frac{1}{N} \sum_i (p_i - c_i) \tag{2}$$

where N is the number of days considered and the set $\{p_i\}_i$ (resp. $\{c_i\}_i$) is the ordered set of the daily expressed mood in passenger tweets (resp. in airline customer service tweets), as for the airline empathy score Ξ presented in equation (1).

The airline sentiment gap Δ goes from -1 to 1 with a gap of 0 indicating that airline customer services and passengers express the same average mood in their tweets. A gap of 1 indicates a mood expressed by an airline customer service equal to 1 (i.e. the highest possible mood) and a mood expressed by the airline passengers equal to 0 (i.e. the lowest possible mood) on every day of the considered period. A gap of -1 indicates the opposite scenario.

Table II shows the ranks and scores of the seven airlines associated with each of the two passenger-centric metrics proposed in this section. Both the empathy score Ξ and the sentiment gap Δ were calculated over the period from March 1st 2020 to March 31st 2020.

III. Keyword-based metrics

A. Cancellations

When some exceptional situation occurs, an important increase in the use of specific keywords within the stream of tweets written by the affected users can take place. For example, if many cancellations occur, many passengers will connect to Twitter and write tweets containing the keyword "cancel" to express their concerns directly to the airline they have bought tickets from. In this analysis, any word starting with the keyword "cancel", such as "cancellation" or "cancelled", is considered as a keyword "cancel".

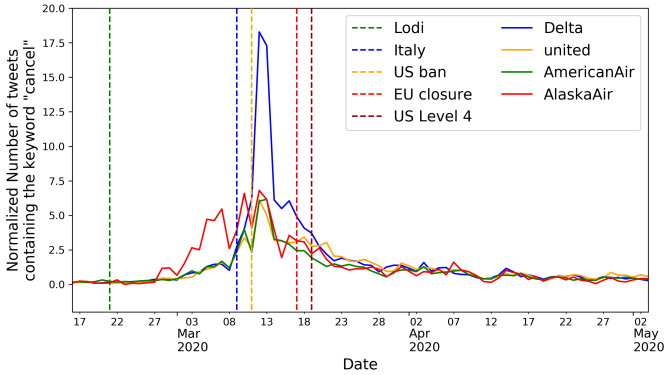
Figure 5 shows the evolution of the normalized number of tweets written by passengers and containing the keyword "cancel" between February 16th 2020 and May 3rd 2020 for four US legacy airlines and four US low-cost carriers. The normalization is based on the total number of passengers carried by each airline in 2018 and available in the yearly BTS reports [62].

Figure 5(a) indicates that the passengers of the four legacy airlines react as early as Italy's lockdown announcement with an important increase in the number of tweets containing the keyword "cancel". A second spike in the number of passenger tweets containing the keyword "cancel" then occurs once the US announces that it bans all travelers from the EU, China and Iran. Figure 5(a) shows that Delta Air Line passengers were, in proportion, about three times more vocal about cancellations on Twitter than the other legacy airlines at this period. This could be an indication that Delta Air Line had a greater proportion of passengers traveling within or through the EU at that time. The number of tweets from Alaska Airlines passengers containing the keyword "cancel" had an early spike compared to the tweets written by passengers from the other legacy airlines. That early spike could be linked to the fact that most of the early US cases of COVID-19 were discovered on the US West Coast first, which is where the main hub of Alaska Airlines is located.

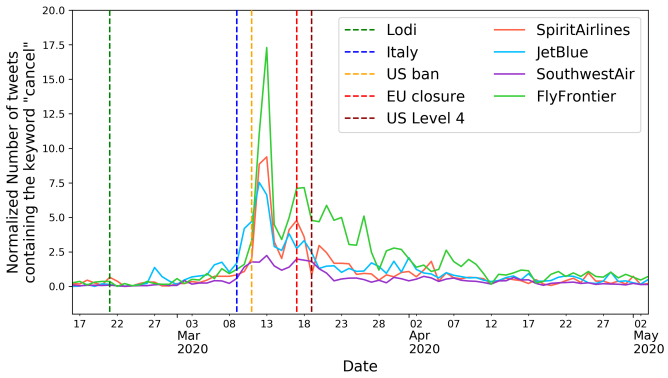
Figure 5(b) shows the evolution of the number of tweets containing the keyword "cancel" written by passengers of the four low-cost carriers. Southwest Airlines passengers were, in proportion, less vocal on Twitter on the matter of cancellation than passengers of the other low-cost carriers, with a slight increase in the number of tweets containing the keyword "cancel" that is almost entirely contained within the period between the announcement of Italy's lockdown and the start

Table II: Airline ranking based on the proposed empathy score Ξ and the sentiment gap Δ applied to the period of March 1st 2020 to March 31st 2020.

Rank	Airline	Empathy Score	Rank	Airline	Sentiment Gap
1	Alaska Airlines	0.476	1	Frontier Airlines	0.104
2	Southwest Airlines	0.456	2	Alaska Airlines	0.179
3	Frontier Airlines	0.374	3	Delta Air Lines	0.228
4	Spirit Airlines	0.146	4	JetBlue Airways	0.228
5	United Airlines	0.129	5	Spirit Airlines	0.237
6	JetBlue Airways	0.066	6	Southwest Airlines	0.244
7	Delta Air Lines	0.029	7	United Airlines	0.246
8	American Airlines	-0.393	8	American Airlines	0.260



(a) From passengers of legacy airlines

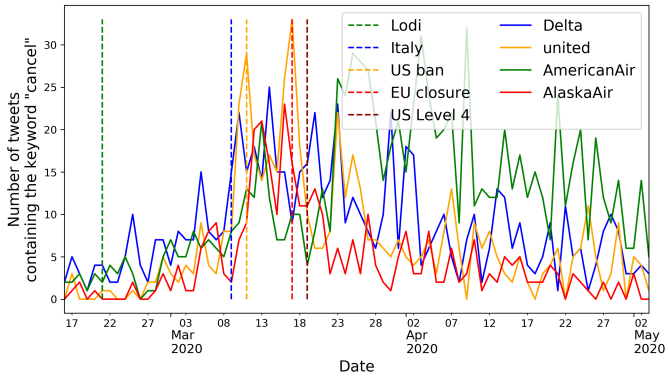


(b) From passengers of low-cost airlines

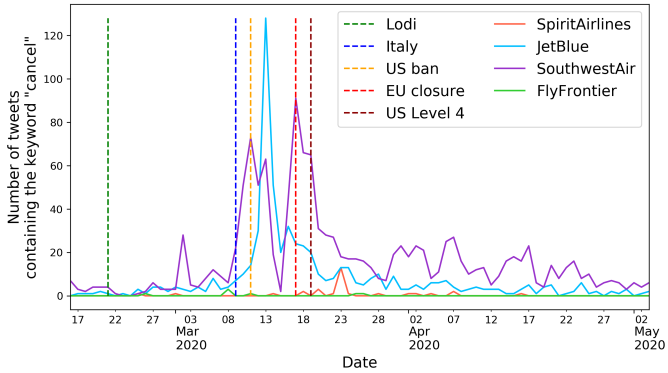
Figure 5: Number of tweets containing the keyword "cancel" and written by passengers normalized by the number of transported passengers per carrier over the year 2018 using BTS data [62]

of the US Level 4 Global Health Travel Advisory. JetBlue Airways passengers display a behavior similar to passengers of legacy airlines in this case. Passengers of Spirit Airlines and Frontier Airlines waited until the US travel ban announcement to communicate massively on Twitter their concerns using the word "cancel". The second spike in the number of tweets containing the keyword "cancel" starting at the announcement of the EU border closure is more important and lasts longer for tweets written by passengers of Frontier Airlines.

Figure 6 shows the evolution of the number of tweets containing the keyword "cancel" and written by airline customer services between February 16th 2020 and May 3rd 2020 for the same four US legacy airlines and three US low-cost carriers. Please note that the y-axis scale is different in Figure 6(a) and Figure 6(b).



(a) From customer service of legacy airlines



(b) From customer service of low-cost airlines

Figure 6: Number of tweets containing the keyword "cancel" in tweets written by airline customer services

Regarding tweets written by legacy airline customer services, the evolution of the number of tweets containing the keyword "cancel" shown in Figure 6(a) presents similarities for three of the four airlines. There is a significant increase in the number of customer service tweets containing the keyword "cancel" starting the day Italy announced its lockdown and then a slow decrease. For tweets written by American Airlines customer service, the number of tweets containing the keyword "cancel" increases as for the other three airlines, but it does not decrease afterwards but fluctuates at a level more important

than during the period before the travel restriction measures where announced.

Regarding low-cost carriers, Figure 6(b) shows that each carrier use the keyword "cancel" on different occasions. The number of occurrences of the keyword "cancel" within tweets written by Southwest Airlines passengers has two important spikes around each of the US announcements referenced in the plot. JetBlue has a single massive spike on March 13th 2020. Both carriers then spent more than two weeks with a higher level of occurrences of the keyword "cancel" than in February 2020. Spirit Airlines customer service never wrote more than three tweets containing the keyword "cancel" in a day except on March 23rd 2020. Frontier Airlines customer service used the keyword "cancel" only in six tweets over the full month of March 2020.

Based on the observations from the plots in Figure 5, an important increase in the normalized number of passenger tweets containing the keyword "cancel" can be treated as an unwanted situation that airlines have to deal with.

Definition 1: A keyword-related **Twitter situation** is defined as an increase over a predefined threshold of the normalized number of passenger-written tweets containing the keyword.

Two metrics to measure the airline reaction to such a situation are proposed here. The aim of the first metric is to measure the effectiveness of the airline response to these keyword-related situations.

Proposed passenger-centric metric 3: The keyword-related Twitter situation **quality response** score of an airline is the time needed for the airline to bring the normalized number of passenger tweets containing the keyword below a predefined threshold.

The Twitter situation quality response score associated to the keyword "cancel" with a threshold of q normalized tweets κ^q_{cancel} is calculated using the following formula:

$$\kappa^q_{\text{cancel}} = d^q_{f,\text{cancel}} - d^q_{0,\text{cancel}} \tag{3}$$

where $d^q_{0,\text{cancel}}$ is defined as the first day of the considered period where the normalized number of passenger tweets containing the keyword "cancel" is greater than q, and $d^q_{f,\text{cancel}}$ is defined as the last day of the considered period where the normalized number of passenger tweets containing the keyword "cancel" is greater than q.

This proposed quality metric measures the time needed for the airline to bring the number of passenger tweets containing the keyword back to a normal state. When measuring the response of long term perturbations, such as the COVID pandemic, this time is measured in days.

The number of passenger tweets containing the keyword is normalized by the total number of passengers carried by the airline over the year 2018 in this case, similarly to the data presented in Figure 5, and this normalization should be updated with the most recent numbers once they are available.

The aim of the second metric is to measure the communication effort produced by the airline in order to handle the situation linked to the increase of number of tweets containing the keyword under consideration.

Proposed passenger-centric metric 4: The keyword-related Twitter situation **quantity response** score of an airline is calculated by integrating the number of tweets containing the keyword and written by the airline customer service over the number of days associated to the keyword-related Twitter situation.

The formula used to calculate the Twitter situation quality response score associated to the keyword "cancel" with a threshold of q normalized tweets γ^q_{cancel} is the following:

$$\gamma^q_{\text{cancel}} = \int_{d^q_{0,\text{cancel}}}^{d^q_{f,\text{cancel}}} n_{\text{cancel}}(t)\ dt \tag{4}$$

where $d^q_{0,\text{cancel}}$ and $d^q_{0,\text{cancel}}$ are the same as for the quality response score κ^q_{cancel} in equation (3), and $n_{\text{cancel}}(t)$ is the number of tweets written by the airline customer service containing the keyword "cancel" on day t.

Table III presents these two proposed metrics in the case of the keyword "cancel" considering that the predefined threshold indicating when a situation starts and ends is 1. Table III illustrates the necessity of considering both the quality response score and the quantity response score hand in hand. Southwest Airlines has the best scores from both perspective but Spirit Airlines has the second best quality response score but the second worst quantity response score. This would indicate that passengers from Spirit Airlines are more resilient to cancellation situations than passengers of the other airlines: They go back to a close-to normal Twitter chatter about cancellation with almost no cancellation related communication efforts on Twitter of Spirit Airlines.

B. Refund

Figure 7 shows the evolution of the normalized number of tweets containing the keyword "refund" and written by passengers from February 16th 2020 to May 3rd 2020 for the same eight US airlines using the same normalization process as for the keyword "cancel".

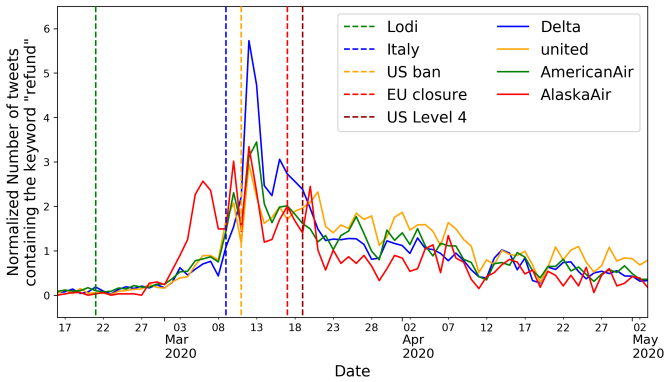
The evolution of the number of passenger tweets containing the keyword "refund" is similar to the evolution of the number of occurrences of the keyword "cancel" but at a lower proportion. Figure 7(a) shows that the number of occurrences of the keyword "refund" in tweets written by passengers of all four legacy airlines steeply increases at the announcement of Italy's lockdown and then very slowly decreases. Passengers of Alaska Airlines have an anticipated spike in the number of tweets containing the keyword "refund" at the beginning of March 2020. Figure 7(b) shows that the increase in the number of tweets containing the keyword "refund" and written by Southwest Airlines passengers is still lower than the number of tweets containing the keyword "refund" and written by the passengers of the other low-cost carriers. The number of tweets containing the keyword "refund" and written by Southwest Airlines passengers gets back to a normal level faster than for the passengers of the other low-cost carriers. The spike in the number of tweets containing the keyword "refund" and written by Spirit Airlines and Frontier Airlines passengers starts only at the announcement of the US travel ban.

Table III: Airline ranking based on the ”cancel”-related Twitter situation quality and quantity response scores applied to the period of March 1st 2020 to April 30th 2020.

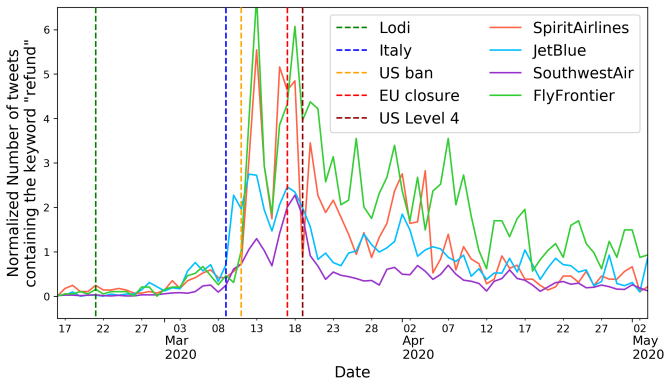
Rank	Airline	Quality (days)	Rank	Airline	Quantity
1	Southwest Airlines	11	1	Southwest Airlines	50.64
2	Spirit Airlines	26	2	American Airlines	15.34
3	United Airlines	34	3	Delta Air Lines	11.98
4	American Airlines	35	4	United Airlines	11.62
5	Delta Air Lines	41	5	JetBlue Airways	10.28
6	Alaska Airlines	44	6	Alaska Airlines	6.82
7	Frontier Airlines	53	7	Spirit Airlines	0.96
8	JetBlue Airways	54	8	Frontier Airlines	0.11

Table IV: Airline ranking based on the ”refund”-related Twitter situation quality and quantity scores applied to the period of March 1st 2020 to April 30th 2020.

Rank	Airline	Quality (days)	Rank	Airline	Quantity
1	Southwest Airlines	8	1	Southwest Airlines	32.25
2	Spirit Airlines	29	2	American Airlines	22.81
3	American Airlines	31	3	United Airlines	7.67
4	Alaska Airlines	35	4	Delta Air Lines	7.46
5	Delta Air Lines	37	5	Alaska Airlines	4.03
6	JetBlue Airways	39	6	JetBlue Airways	3.03
7	United Airlines	51	7	Frontier Airlines	0.02
7	Frontier Airlines	51	8	Spirit Airlines	0.00



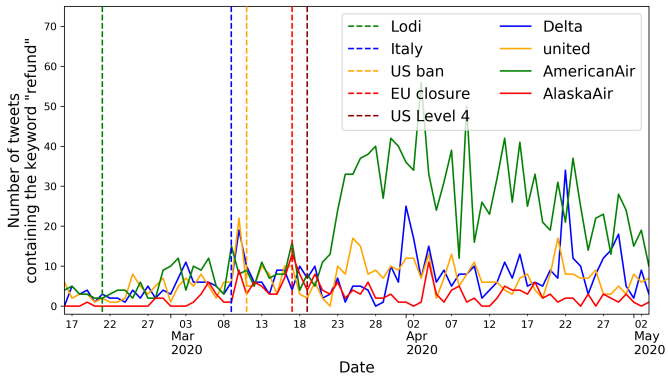
(a) From passengers of legacy airlines



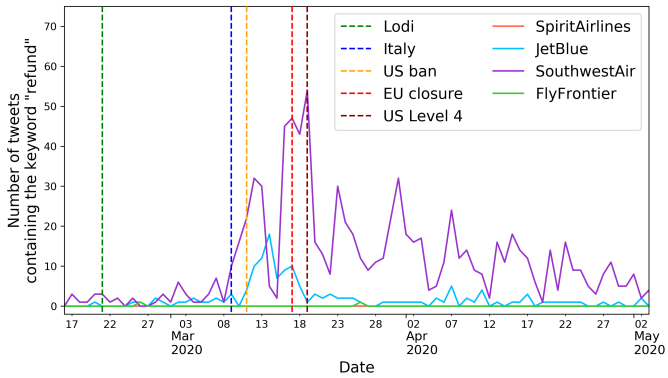
(b) From passengers of low-cost airlines

Figure 7: Number of tweets containing the keyword ”refund” and written by passengers normalized by the number of transported passengers per carrier over the year 2018 using BTS data [62]

Figure 8 shows the evolution of the number of tweets containing the keyword ”refund” and written by airline customer services from February 16th 2020 to May 3rd 2020 for the same eight US airlines.



(a) From customer service of legacy airlines



(b) From customer service of low-cost airlines

Figure 8: Number of tweets containing the keyword ”refund” and written by airline customer services

Figure 8(a) shows the evolution of the number of tweets

containing the keyword "refund" and written by the customer services of the four considered legacy airlines. The initial increase is similar than for the keyword "cancel" (Figure 6(a)), however there is then a second increase towards the end of March 2020, this increase being most visible within the tweets written by American Airlines customer service. From a low-cost carrier perspective, Figure 8(b) illustrates the same characteristics as in Figure 6(b): There are two spikes around the US announcements for the number of tweets containing the keyword "refund" in tweets written by Southwest Airlines customer service, this time with higher fluctuations afterwards, and one major spike on March 13th 2020 for the number of tweets containing the keyword "refund" and written by JetBlue Airways customer service. Only one tweet containing the keyword "refund" was written by Frontier Airlines customer service over the month of March 2020 and none written by Spirit Airlines customer service since February 16th 2020.

The same two metrics associated to the "cancel"-related Twitter situation presented in Section III-A, i.e. the quality response score and the quantity response score, can be used for this "refund"-related Twitter situation. Table IV presents these two proposed metrics in the case of the keyword "refund" using the same predefined threshold of 1 for delimiting a Twitter situation.

As for the handling of the "cancel"-related Twitter situation, Southwest Airlines had the most effective (best quality response score) and most proactive (best quantity response score) of the eight airlines. The same resilience is shown by passengers of Spirit Airlines during this "Refund"-related Twitter situation as for the "cancel"-related Twitter situation.

IV. Discussion & Conclusion

A. Score summary

Figure 9 presents a radar plot for each of the eight considered airlines indicating their normalized scores.

The normalizations were conducted using the following formulas:

$$\hat{\Xi} = \frac{1+\Xi}{2} \tag{5}$$

$$\hat{\Delta} = \frac{1-\Delta}{2} \tag{6}$$

$$\hat{\kappa}^q_{\text{keyword}} = \frac{1-\kappa^q_{\text{keyword}}}{\delta T} \tag{7}$$

$$\hat{\gamma}^q_{\text{keyword}} = \frac{\gamma^q_{\text{keyword}}}{\delta T} \tag{8}$$

where δT is the number of days of the full period over which the keyword-related Twitter situation response scores are calculated. All-but-one of the normalized scores go from the worst score of 0 to a good score of 1. The score can be greater than 1 in the case of a keyword-related Twitter situation response quantity score, but that scenario did not occur here. Regarding the normalized sentiment gap, a score of 0.5 indicates a normal score of 0, a normalized score of 0 indicates a score of 1 and a normalized score of 1 indicates a score of -1.

B. Discussion

As can be seen in Figure 9, each airline has its own "Twitter profile". Passengers are then free to integrate these different profiles in their decision process for choosing the airline that corresponds the most to their travel needs and wants. Traditionally, the airline and airport choices are shown to be based on fare, access time and journey time [63], [64]. These studies do not take the airlines reputation among passengers as a decision parameter, and the proposed metrics could provide an additional decision layer for passengers.

For example, some risk-averse passengers could decide to opt for an airline that has better "refund"-related scores if they prefer a refund when flights are cancelled, rather than choosing an airline with a lower fare. Similarly, some passengers can consider that the flight experience is important in their airline decision and use the empathy and sentiment gap scores to help them decide which airline choose. After their experience with the airline, passengers can tweet about it, which will then be taken into account in the next score update. This process corresponds to a feedback loop illustrated in Figure 10.

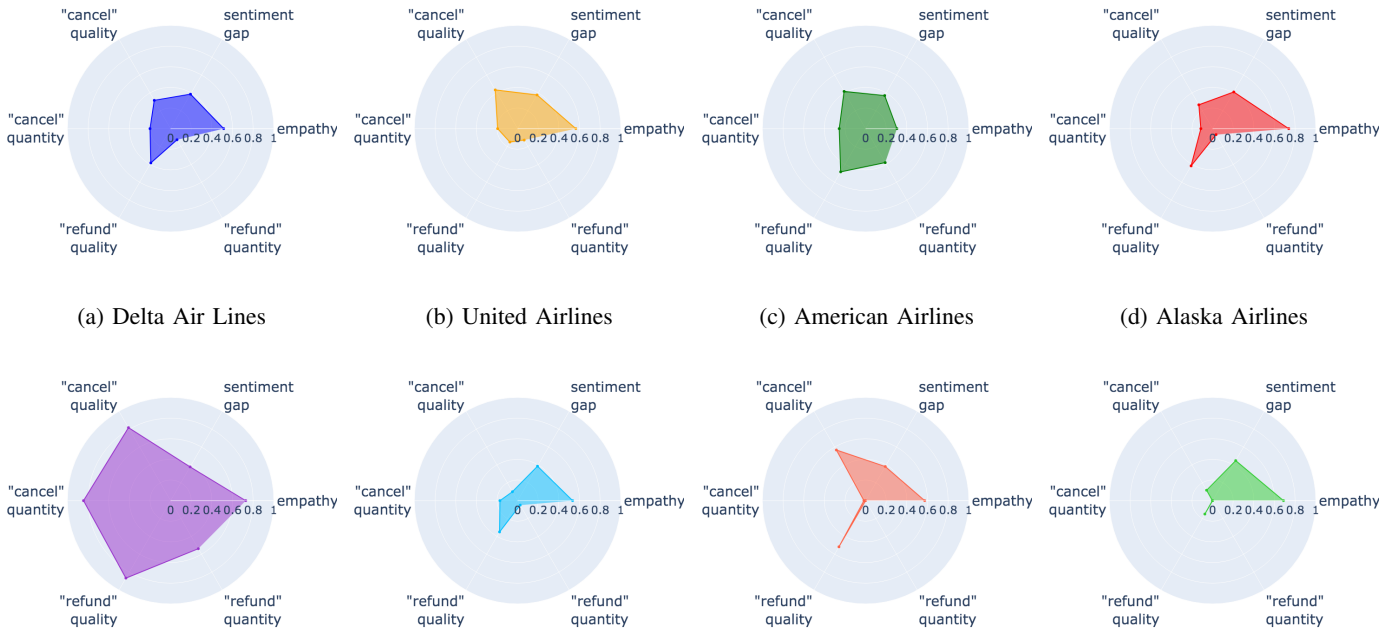
On the other hand, airlines can also compare their Twitter profiles provided in Figure 9 in order to improve their interactions with their passengers. For example, an airline with a clear description of their cancellation procedures on their website could use the "cancel" and "refund" related scores to verify if this information is actually easily accessible to passengers and if adequate communication is made on its availability. For example, a low "cancel" quality score would indicate that passengers already have access to the cancellation information. The proposed metrics can therefore also be used as a part of a feedback loop for airlines, regarding how their policies are implemented and if changes are necessary. Such a feedback loop is illustrated in Figure 11.

The proposed feedback loops are based on unsolicited feedback, which is usually biased towards extreme negative and extreme positive feedback. The bias potentially remaining in the metrics can be corrected with the collaboration of airlines thanks to the access to their data about their passenger experience surveys.

C. Conclusion

The proposed passenger-centric metrics were built using Twitter data, which have the major advantage of being available in real-time, and can therefore be easily updated on an hourly basis if needed. Discussion between federal agencies, airlines and passengers should be undertaken in order to further tune the proposed metrics in order to meet the expectations of all concerned parties.

The proposed metrics have thee added benefit of enabling each passenger and airline to actively influence the scores. It should however be emphasized here that the metrics measure essentially the communication quality and quantity between airlines and passengers via Twitter, and should therefore still be complemented with traditional flight-centric measures for completeness.



(a) Delta Air Lines (b) United Airlines (c) American Airlines (d) Alaska Airlines

(e) Southwest Airlines (f) JetBlue Airways (g) Spirit Airlines (h) Frontier Airlines

Figure 9: Radar plots of the normalized scores associated to the proposed passenger-centric metrics

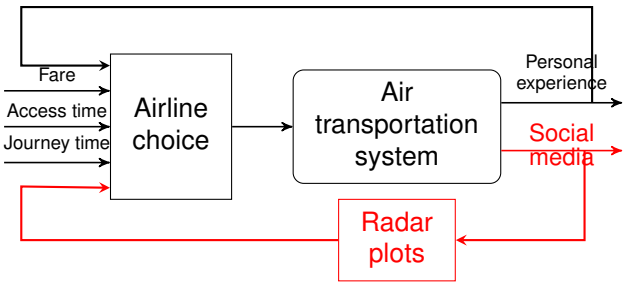


Figure 10: Proposed feedback loop for passengers

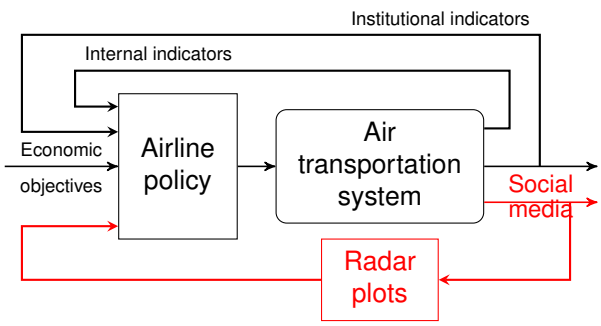


Figure 11: Proposed feedback loop for airlines

This study focused on the effects of the travel restriction measures linked to a major disruption taking its course over an important number of days and tailored the proposed metrics for this timespan. Future studies could also investigate into the adaptation of some of these proposed passenger-centric metrics to measure effects on a smaller scale, e.g. over a single day or a few hours.

Acknowledgments

The authors would like to thank Nikunj Oza from NASA-Ames, the French *École Nationale de l'Aviation Civile* and the King Abdullah University of Science and Technology for their financial support.

The authors would also like to deeply thank all workers and researchers associated in the fight against the COVID-19 pandemic, with a special thought to health-care workers and providers.

References

[1] New York Times, "Coronavirus Travel Restrictions, Across the Globe," 2020. [Online]. Available: https://www.nytimes.com/article/coronavirus-travel-restrictions.html

[2] WorldAtlas, "Which countries are in mandatory lockdown due to COVID-19?" 2020. [Online]. Available: https://www.worldatlas.com/which-countries-are-in-mandatory-lockdown-due-to-covid-19.html

[3] US Department of State, "Global Level 4 Health Advisory - Do Not Travel," 2020. [Online]. Available: https://travel.state.gov/content/travel/en/traveladvisories/ea/travel-advisory-alert-global-level-4-health-advisory-issue.html

[4] United States Customs and Border Protection, "Airport wait times," 2020. [Online]. Available: {https://awt.cbp.gov}

[5] P. Monmousseau, A. Marzuoli, C. Bosson, E. Feron, and D. Delahaye, "Doorway to the United States: An Exploration of Customs and Border Protection Data," in *38th Digital Avionics Systems Conference*, San Diego, California, USA, 2019.

[6] S. Namilae, A. Srinivasan, A. Mubayi, M. Scotch, and R. Pahle, "Self-propelled pedestrian dynamics model: Application to passenger movement and infection propagation in airplanes," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 248–260, Jan. 2017.

[7] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, p. eaba9757, Mar. 2020.

[8] J. B. Sobieralski, "COVID-19 and airline employment: Insights from historical uncertainty shocks to the industry," *Transportation Research Interdisciplinary Perspectives*, vol. 5, p. 100123, May 2020.

[9] Bureau of Transportation Statistics, "Bureau of Transportation Statistics, About BTS," 2018. [Online]. Available: http://www.rita.dot.gov/bts/about

[10] S. Truxal, "COVID-19 airport slot rules: What's changed and what's next for European airlines?" 2020.

[11] Business Insider, "Airlines are burning thousands of gallons of fuel flying empty 'ghost' planes so they can keep their flight slots during the coronavirus outbreak," 2020. [Online]. Available: https://www.businessinsider.fr/us/coronavirus-airlines-run-empty-ghost-flights-planes-passengers-outbreak-covid-2020-3

[12] Congress of the United States of America, "Coronavirus Aid, Relief, and Economic Security (CARES) Act (H.R. 748, Public Law 116-136)," 2020. [Online]. Available: https://www.congress.gov/116/bills/hr748/BILLS-116hr748enr.pdf

[13] S. Bratu and C. Barnhart, "An Analysis of Passenger Delays Using Flight Operations and Passenger Booking Data," *Air Traffic Control Quarterly*, vol. 13, no. 1, pp. 1–27, 2005.

[14] ——, "Flight operations recovery: New approaches considering passenger recovery," *Journal of Scheduling*, vol. 9, no. 3, pp. 279–298, Jun. 2006.

[15] D. Wang, D. L. Sherry, and D. G. Donohue, "Passenger Trip Time Metric for Air Transportation," in *The 2nd International Conference on Research in Air Transportation*, 2006.

[16] D. Wang, "Methods for analysis of passenger trip performance in a complex networked transportation system," Doctor of Philosophy, George Mason University, Fairfax, Virginia, USA, 2007.

[17] Y. O. Gawdiak and T. Diana, "NextGen Metrics for the Joint Planning and Development Office," vol. 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, including the AIAA Balloon Systems Conference and 19th AIAA Lighter-Than, 2011.

[18] M. Darecki, C. Edelstenne, E. Fernandez, P. Hartman, J.-P. Herteman, M. Kerkloch, I. King, P. Ky, M. Mathieu, G. Orsi, G. Schotman, C. Smith, and J.-D. Wörner, *Flightpath 2050: Europe's Vision for Aviation ; Maintaining Global Leadership and Serving Society's Needs ; Report of the High-Level Group on Aviation Research*, E. Commission, Ed. Luxembourg: European Commission, 2011.

[19] A. Cook, G. Tanner, S. Cristóbal, and M. Zanin, "Passenger-Oriented Enhanced Metrics," 2012.

[20] I. Laplace, A. Marzuoli, and E. Feron, "META-CDM: Multimodal, Efficient Transportation in Airports and Collaborative Decision Making," in *Airports in Urban Networks 2014 (AUN2014)*, Paris, 2014.

[21] S. H. Kim, A. Marzuoli, J.-P. Clarke, D. Delahaye, and E. Feron, "Airport Gate Scheduling for Passengers, Aircraft, and Operation," in *Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)*, Chicago, Illinois, Jun. 2013.

[22] L. Dray, A. Marzuoli, and A. Evans, "Air Transportation and Multimodal, Collaborative Decision Making during Adverse Events," in *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*, Lisboa, Portugal, Jun. 2015.

[23] EUROCONTROL and Federal Aviation Administration Air Traffic Organization System Operations Services, "2015 Comparison of Air Traffic Management-Related Operational Performance: U.S./Europe," Tech. Rep., Aug. 2016.

[24] A. C. Lemer, "Measuring performance of airport passenger terminals," *Transportation Research Part A: Policy and Practice*, vol. 26, no. 1, pp. 37–45, Jan. 1992.

[25] S.-H. Tsaur, T.-Y. Chang, and C.-H. Yen, "The evaluation of airline service quality by fuzzy MCDM," *Tourism Management*, vol. 23, no. 2, pp. 107–115, Apr. 2002.

[26] J. A. Hunter, "A correlational study of how airline customer service and consumer perception of airline customer service affect the air rage phenomenon," *Journal of Air Transportation*, vol. 11, no. 3, pp. 78–109, 2006.

[27] J. de Oña and R. de Oña, "Quality of Service in Public Transport Based on Customer Satisfaction Surveys: A Review and Assessment of Methodological Approaches," *Transportation Science*, vol. 49, no. 3, pp. 605–622, Aug. 2015.

[28] F. Pakdil and Ö. Aydın, "Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores," *Journal of Air Transport Management*, vol. 13, no. 4, pp. 229–237, Jul. 2007.

[29] H. Nikoue, A. Marzuoli, J.-P. Clarke, E. Feron, and J. Peters, "Passenger Flow Predictions at Sydney International Airport: A Data-Driven Queuing Approach," *arXiv:1508.04839 [cs]*, Aug. 2015.

[30] W. Huang, Y. Lin, B. Lin, and L. Zhao, "Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data," *Energy & Buildings*, vol. 203, p. 19, 2019.

[31] J. Van den Heuvel, D. Ton, and K. Hermansen, "Advances in measuring pedestrians at dutch train stations using bluetooth, wifi and infrared technology," in *Traffic and Granular Flow'15*. Springer, 2016, pp. 11–18.

[32] A. Marzuoli, E. Boidot, E. Feron, and A. Srivastava, "Implementing and Validating Air Passenger–Centric Metrics Using Mobile Phone Data," *Journal of Aerospace Information Systems*, vol. 16, no. 4, pp. 132–147, Apr. 2019.

[33] A. Marzuoli, P. Monmousseau, and E. Feron, "Passenger-centric metrics for Air Transportation leveraging mobile phone and Twitter data," in *Data-Driven Intelligent Transportation Workshop - IEEE International Conference on Data Mining 2018*, Singapore, Nov. 2018.

[34] P. García-Albertos, O. G. Cantú Ros, R. Herranz, and C. Ciruelos, "Understanding Door-to-Door Travel Times from Opportunistically Collected Mobile Phone Records," in *SESAR Innovation Days 2017*, 2017.

[35] A. Pruyn and A. Smidts, "Effects of waiting on the satisfaction with the service: Beyond objective time measures," *International Journal of Research in Marketing*, vol. 15, no. 4, pp. 321–334, Oct. 1998.

[36] K. E. Watkins, B. Ferris, A. Borning, G. S. Rutherford, and D. Layton, "Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 8, pp. 839–848, Oct. 2011.

[37] S. E. Sampson, "Ramifications of Monitoring Service Quality Through Passively Solicited Customer Feedback," *Decision Sciences*, vol. 27, no. 4, pp. 601–622, Dec. 1996.

[38] A. S. Abrahams, J. Jiao, W. Fan, G. A. Wang, and Z. Zhang, "What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings," *Decision Support Systems*, vol. 55, no. 4, pp. 871–882, Nov. 2013.

[39] G. Kane, "Reimagining customer service at klm using facebook and twitter," *MIT Sloan Management Review*, vol. 55, no. 4, pp. 1–6, 2014.

[40] L. Palen, K. Starbird, S. Vieweg, and A. Hughes, "Twitter-based information distribution during the 2009 Red River Valley flood threat," *Bulletin of the American Society for Information Science and Technology*, vol. 36, no. 5, pp. 13–17, Jun. 2010.

[41] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," p. 10, 2010.

[42] K. Kireyev, L. Palen, and K. M. Anderson, "Applications of Topics Models to Analysis of Disaster-Related Twitter Data," p. 4, 2009.

[43] T. Terpstra and R. Stronkman, "Towards a realtime Twitter analysis during crises for operational crisis management," p. 10, 2012.

[44] S. Priya, M. Bhanu, S. K. Dandapat, K. Ghosh, and J. Chandra, "TAQE: Tweet Retrieval-Based Infrastructure Damage Assessment During Disasters," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 389–403, Apr. 2020.

[45] H. Srivastava and R. Sankar, "Information Dissemination From Social Network for Extreme Weather Scenario," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 319–328, Apr. 2020.

[46] J. O. Breen, "Mining twitter for airline consumer sentiment," *Practical text mining and statistical analysis for non-structured text data applications*, vol. 133, 2012.

[47] Y. Wan and Q. Gao, "An ensemble sentiment classification system of Twitter data for airline services analysis," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. Atlantic City, NJ, USA: IEEE, Nov. 2015, pp. 1318–1325.

[48] K. Siau, "An Approach to Sentiment Analysis – The Case of Airline Quality Rating," *Pacific Asia Conference on Information Systems (PACIS 2014)*, 2014.

[49] B. D. Bowen, D. E. Headley, and J. R. Luedtke, *Airline Quality Rating*. Wichita State University, National Institute for Aviation Research, 1991, vol. 91.

[50] F. Misopoulos, M. Mitic, A. Kapoulas, and C. Karapiperis, "Uncovering customer service experiences with Twitter: The case of airline industry," *Management Decision*, vol. 52, no. 4, pp. 705–723, May 2014.

[51] P. Gunarathne, H. Rui, and A. Seidmann, "Customer Service on Social Media: The Effect of Customer Popularity and Sentiment on Airline Response," in *2015 48th Hawaii International Conference on System Sciences*. HI, USA: IEEE, Jan. 2015, pp. 3288–3297.

[52] P. Monmousseau, A. Marzuoli, E. Feron, and D. Delahaye, "Predicting and Analyzing US Air Traffic Delays using Passenger-centric Datasources," in *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)*, Vienna, Austria, 2019.

[53] ——, "Passengers on social media: A real-time estimator of the state of the US air transportation system," in *ENRI Int. Workshop on ATM/CNS (EIWAC 2019)*, Tokyo, Japan, 2019.

[54] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proceedings of the ACL Student Research Workshop on - ACL '05*. Ann Arbor, Michigan: Association for Computational Linguistics, 2005, p. 43.

[55] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *CS224N project report, Stanford*, vol. 1, 2009.

[56] Kaggle, "Twitter US airline sentiment," https://www.kaggle.com/crowdflower/twitter-airline-sentiment, 2018.

[57] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*. Springer, 1995, pp. 23–37.

[58] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[59] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[60] T. F. Chan, G. H. Golub, and R. J. LeVeque, "Updating Formulae and a Pairwise Algorithm for Computing Sample Variances," in *COMPSTAT 1982 5th Symposium Held at Toulouse 1982*, H. Caussinus, P. Ettinger, and R. Tomassone, Eds. Heidelberg: Physica-Verlag HD, 1982, pp. 30–41.

[61] H.-F. Yu, F.-L. Huang, and C.-J. Lin, "Dual coordinate descent methods for logistic regression and maximum entropy models," *Machine Learning*, vol. 85, no. 1-2, pp. 41–75, 2011.

[62] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine Learning in Python," *Machine Learning in Python*, 2011.

[63] Bureau of Transportation Statistics, "2018 Traffic Data for U.S Airlines and Foreign Airlines U.S. Flights," 2020. [Online]. Available: https://www.bts.dot.gov/newsroom/2018-traffic-data-us-airlines-and-foreign-airlines-us-flights

[64] E. Pels, P. Nijkamp, and P. Rietveld, "Access to and competition between airports: A case study for the San Francisco Bay area," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 1, pp. 71–83, Jan. 2003.

[65] S.-Y. Jung and K.-E. Yoo, "Passenger airline choice behavior for domestic short-haul travel in South Korea," *Journal of Air Transport Management*, vol. 38, pp. 43–47, Jun. 2014.