Received April 2, 2019, accepted April 26, 2019, date of publication May 7, 2019, date of current version May 20, 2019.
Digital Object Identifier 10.1109/ACCESS.2019.2915107

Using Twitter to Infer User Satisfaction With Public Transport: The Case of Santiago, Chile

JOSÉ TOMÁS MÉNDEZ[1], (Member, IEEE), HANS LOBEL[1,2], DENIS PARRA[2], AND JUAN CARLOS HERRERA[1]
[1]Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Santiago 7820436 , Chile
[2]Computer Science Department, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile

Corresponding author: José Tomás Méndez (jtmendez@uc.cl)

This work was supported in part by the BRT+ Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF).

ABSTRACT User satisfaction is an important aspect to consider in any public transport system, and as such, regular and sound measurements of its levels are fundamental. However, typical evaluation schemes involve costly and time-consuming surveys. As a consequence, their frequency is not enough to properly and timely characterize the satisfaction of the users. In this paper, we propose a methodology, based on Twitter data, to capture the satisfaction of a large mass of users of public transport, allowing us to improve the characterization and location of their satisfaction level. We analyzed a massive volume of tweets referring to the public transport system in Santiago, Chile (Transantiago) using text mining techniques, such as sentiment analysis and topic modeling, in order to capture and group bus users' expressions. Results show that, although the level of detail and variety of answers obtained from surveys are higher than the ones obtained by our method, the amount of bus stops and bus services covered by the proposed scheme is larger. Moreover, the proposed methodology can be effectively used to diagnose problems in a timely manner, as it is able to identify and locate trends, and issues related to bus operating firms, whereas surveys tend to produce average answers. Based on the consistency and logic of the results, we argue that the proposed methodology can be used as a valuable complement to surveys, as both present different, but compatible characteristics.

INDEX TERMS Natural language processing, public transport, sentiment analysis, topic modeling, user satisfaction.

I. INTRODUCTION

Transantiago is the name given to the public transportation system of Santiago, Chile. It was implemented more than a decade ago on February 2007, and by 2016 includes an underground railway covering 104 kilometers and approximately 6,600 buses distributed in 379 services. More than two million people use the system daily [1].

Due to several issues on its implementation, Transantiago has presented increasing disapproval rates among the entire population, according to a nation-wide survey [2]. Transantiago users' satisfaction has been systematically measured by the corresponding authority by surveying 5,000 users every year.[1] However, the high cost of surveys, in terms of time and money, presents a major challenge, as it strongly limits the ability to timely capture users' opinions, demands and problems with a proper spatial coverage.

An alternative to bypass the limitations of traditional surveys is to use massive sources of free data, such as the ones provided by social networks. Nowadays people reach to them to express their opinions of their day to day activities. In particular, Twitter, due to its instantaneous nature, could serve as a source of instant, massive and free information for local authorities.

A. PAPER CONTRIBUTIONS

The objective of this work is to analyze the use of Twitter as a tool capable of measuring the level of satisfaction of bus users in Transantiago. Four research questions are formulated to achieve this objective: (RQ1) Is Twitter mostly used by Transantiago bus users to express their discontent with the system? (RQ2) Can tweets referring to the system be grouped into topics or themes? (RQ3) Is there a bias in using Twitter to

The associate editor coordinating the review of this manuscript and approving it for publication was Yue Cao.

[1]Until 2015, 15,000 surveys were conducted yearly. Because of budget constraint, this number was reduced to 5,000 since 2016.

measure satisfaction levels of Transantiago bus users? (RQ4) Can satisfaction surveys be replaced by information reported on Twitter?

Our approach to infer user satisfaction is based on three main elements:

- The use of well-established text mining techniques [3] in a novel manner in the challenging context of public transport user satisfaction inference, in order to deal with the unstructured nature of the data generated by Twitter. To the best of our knowledge, this is the first time a study of this kind is performed, where the satisfaction of a large mass of public transport users is inferred without using questionnaires.
- The generation of a control group based on tweets produced by a real-time bus information app. This is a key aspect that allowed us to validate the topics, without resorting to expensive, and sometimes unreliable, expert knowledge.
- An in-depth spatio-temporal analysis of tweets, in order to account for biases in crowd-generated data, and to assess their coverage when compared with a traditional survey method.

B. PAPER OUTLINE

The remainder of this article is organized as follows: Section II reviews relevant previous works. Section III presents the characteristics of the datasets used in this research. Section IV describes the proposed methodology to address the research questions, while Section V presents its results. Finally, Section VI presents the conclusions of this research and Section VII discusses its main limitations and future work.

II. RELATED WORK

Diverse methodologies and tools are used for measuring user satisfaction on a public transportation context. Traditional tools such as surveys and mobile applications have taken the lead on the past, being a 26 question poll the preferred methodology of the authority in charge of Transantiago (*DTPM*) [4]. The main benefit of using a questionnaire is that it allows analysts to quantify and generalize the behavior of a large mass of people. Nevertheless, long questionnaires can influence on answers, obtaining results close to the average [5]. Recently, surveys have been upgraded with the inclusion of technology. For example, the questions have been incorporated in smartphones or tablets [6] [7] to reduce the time and cost of getting the information, generate more precise and reliable data, as well as being eco-friendly.

Mobile applications with real-time information of bus arrivals allow transit users to report situations of their trip to other members of the community. Apps such as Roadify [8], TrafficInfo [9] and TranSapp [10] have been implemented in their respective cities. Although they share certain similarities to our methodology, particularly regarding the fact that crowd-generated data are also used, these works present important differences, such as the scale of the study (≈15K registers theirs, ≈90K registers ours), the use of tweets as data source, the type of data (structured reports vs unstructured text), and specially the depth and variety of the analysis and the tools used. To the best of our knowledge, this is the first time a study of this kind and size is performed.

In order to process and analyze text data coming from a social network such as Twitter, we need tools that are able to structure the unstructured text data and to extract relevant information. In this research, we focused on two such tools, namely sentiment analysis and topic modeling.

Sentiment analysis is the computational study of people's opinions, attitudes and emotions expressed in a text [11]. A pilot plan to gauge Chicago transit riders' sentiments was implemented in 2011 by measuring Twitter feeds using the software SentiStrength [12]. Considering that the database analyzed in this investigation is written in Spanish, the need for a sentiment analysis software adapted to this language is evident. In 2015 SentiStrength was modified to measure the sentiment of tweets written in Spanish, related to Spanish politicians [13].

Topic modeling consists of algorithms that aim to characterize a text, based on the topics it deals with [14]. A topic is defined as a group of words that tend to occur together frequently [15]. Most approaches use Latent Dirichlet Allocation (LDA) or a variant [16]. For LDA-based methods, MALLET [17] and the Stanford Topic Modeling Toolbox (TMT)[2] were the most frequently reported tools used. A critical aspect of topic modeling is defining the number of topics a text is going to be modeled into. Most solutions suggest arduous and tedious iterative approaches. A heuristic approach [18] proposes the use of the rate of perplexity change (RPC) as a function of the number of topics as a suitable selector. Between 2005 and 2013 the online user reviews of different public transport agencies in the United States were studied using LDA topic modeling to identify the dimensions with the greatest impact in public transport users' satisfaction [15].

III. DATASET

The time span covered on this investigation includes data from years 2014 to 2016. To get tweets of those years, a code written in Java that bypasses some limitations of Twitter's Official API was used.[3] The code can search for tweets on Twitter's page and automatically save the tweets publicly available on the Twitter Search browser.

As the georeferencing of tweets is not a default option in Twitter [19], we included in the queries the words "transantiago" and the codes of the 11,340 bus stops or the number of the 379 bus services. Based on this, two databases were generated: (i) Stops database, which includes the bus stops' codes and initially summed up to 31,910 tweets, and (ii) Buses database, which includes the bus service's number and initially summed up to 79,999 tweets.

[2] https://nlp.stanford.edu/software/tmt/tmt-0.4/.

[3] https://github.com/Jefferson-Henrique/GetOldTweets-java.

In order to filter the database, all letters were modified to lowercase, any strange characters such as accent marks or the letter ''ñ'' (typically used in Spanish) were changed, stop words were eliminated and tweets from accounts of non-users, such as media or authorities, were removed.

During the filtering process, we identified a significant number of tweets with the same structure, related to a real time bus information app. These tweets were voluntarily posted by users of the app when the public transport agency did not update their bus frequency databases (used as an input for the app), causing it to fail. These tweets were interpreted as a complaint from users to the authorities, therefore they were kept on both databases. Finally, we only included in the datasets tweets generated during workdays. After the cleansing and filtering process, the Stops database included 26,318 tweets and the Buses database included 64,868 tweets.

IV. METHODOLOGY

The following section briefly explains the methodology used to answer the four proposed research questions.

To answer the first question (RQ1), a sentiment analysis of both databases was done. The Spanish version of SentiStrength was not used for the task because its level of accuracy was unreliable for the proposed analysis of this investigation [13]. Therefore, the sentiment analysis consisted of reading and manually classifying (as positive, neutral or negative) approximately ten percent of both databases, adding up to 9,000 tweets combined. Comparing the sentiments identified by the Spanish version of SentiStrength with the manual classification of the 9,000 tweets, only 41% of the tweets were correctly classified by the software.

To answer RQ2, topic modeling techniques are applied to both databases. For this purpose, we grouped the tweets from each database into longer documents, as current techniques show problems when dealing with short documents, such as single tweets. The Stops database was grouped into 34 documents, each containing all the tweets from a certain commune of Santiago.[4] The Buses database was grouped into 28 documents, each containing all the tweets from a type of bus service available in Transantiago. A bus service was classified by its operating firm (seven firms), length (long or short) and predominant direction (north-south or east-west).

To select an adequate number of topics, we applied the RPC methodology [18] by measuring the perplexity of models with different number of topics. Based on this analysis, the appropriate number of topics for the Stops and Buses databases was 37 and 13 topics, respectively. After this, a topic model was generated for each collection of documents (one for the 34 documents of the Stops database and one for the 28 documents of Buses database) using the optimal number of topics in MALLET.

To check and validate topic models [14], we tried to avoid the reliance on expert knowledge [20] by developing an accuracy index for each database. We took advantage of the known structure of tweets coming from the real time bus information app previously identified, and the fact that MALLET recognized the keywords of these tweets as a topic. Thus, by comparing the average percentage of occurrence the model assigned to that topic, with the fraction of the databases represented by those tweets, which is known, we can quantify each model's accuracy.

Regarding RQ3, in order to find any inherent biases or trends in the information obtained from Twitter, we performed a validation analysis based on three different aspects, namely temporal, spatial and operational.

Finally, to answer RQ4, we performed a comparative analysis between the results generated by answering the previous research questions, and the satisfaction surveys currently used by the DTPM. The satisfaction surveys database totaled 27,473 polls (excluding those done on weekends). Each survey has a stop and bus service code; therefore, the same database was used to compare the Stops database and Buses database with tweets. The spatial and user coverage of both methodologies was also compared, as well as their sentiment analysis, topic distribution and predominant topics.

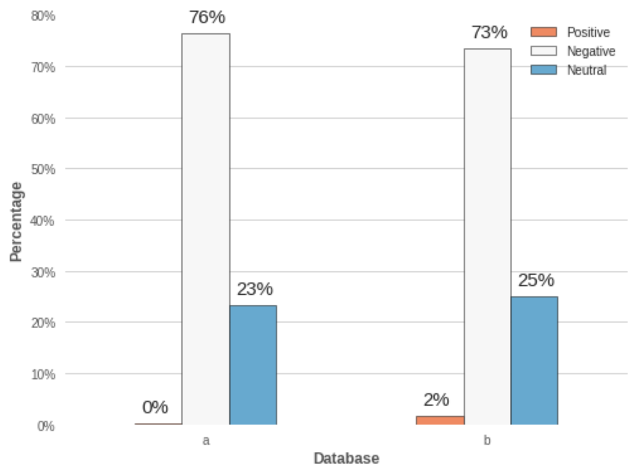


FIGURE 1. Sentiment analysis of tweets, a) Stops database b) Buses database.

V. RESULTS AND DISCUSSION

RQ1: Is Twitter mostly used by Transantiago bus users to express their discontent with the system?

The results of reading and classifying more the 9,000 tweets are presented in Figure 1.

In both databases, approximately 75 percent of tweets express a negative sentiment. It is worth noting that tweets related to the previously identified real-time information app were classified as negative. Most of the remaining percentage are neutral messages, which are mostly information queries. The positive tweets in both databases are quite low, so they can be considered negligible.

Based on these results, we conclude that, as expected, public transport buses users in Santiago use Twitter mainly to complain about the service.

[4] A commune is the smallest administrative subdivision in Chile.

TABLE 1. Topics, Keywords and Examples, Stops Database.

Topic	Key words	Examples
Frequency	Minute, passage, wait, bus, more, since, half, hour.	@subus_ what is happening with the buses 201-223 and 230 ?? Stop PB198 ... 20 Minutes waiting, your frequency is so bad! @Transantiago
App	For, service, stop, information, happens, alsaciaexpress.	@Transantiago There is no information for service B17 in stop PB1563. What's going on?
Information	Arrive, good, how, service, thanks, day, bus, much.	@Transantiago in how much time does g22 arrive to stop PG1407 … Thanks
Stop	"Plates", stop, license plate, bus stop, bus driver, service.	@Transantiago let me inform you that the buses with license plate CJRP-26 and CJRK-90 did not stop at the PG1410 bus stop
Out of order	"Plates", lights, in-transit, stops, supervise, off, signboard.	Ppu bjfk47 line 210 "out of order" by PI339. There are always many "out of order" for this stop. @Transantiago @subus_
Driver	Parked, stop, race, before, skipping, ahead, passengers, leave.	@Transantiago the 225 should stop at the stop PC144 ??? It just parked half a block away and did not pick up passengers.
Overcrowding	Collapsed, full, people, queue, level, saturated, long.	@Transantiago pc614 the stop is collapsed and the c01 are full, send larger buses!
Social	Swarmapp.com, http, Santiago, checkin, Metropolitan, region, square, sq.com.	To the gym (@Stio PC177 [Noruega / Av. Apoquindo] in Santiago, Metropolitan Region) https://www. swarmapp.com/c/
Cleanliness	Taxi, clean, wish, worthy, leave, toilet, infection, focus.	The stop PC128 smells like excrement for WEEKS. Please clean up
Regularity	Predictor, application, arrived, regularize, follow-up, while, empty, just.	@Transantiago Could you regularize the frequency of service 506 and 506e? They always arrive together at the PI167 stop and then you've got to wait for a long time.

TABLE 2. Topics, Keywords and Examples, Buses Database.

Topic	Key words	Examples
Frequency	Minutes, wait, service, more, bus, hour.	I've been waiting 30 minutes for a 428 @Transantiago a disgusting service. Stop Mall Arauco Maipu Pi237
Stop	"Plates", stop, bus stop, license plate, service, bus, bus driver.	@Transantiago the bus with my1938 service c07 did not stop at pc740 I demand that they make it return.
Information	How, bus stop, where, passage, service, arrive, bus.	@Transantiago hello Transantiago in how many minutes the 113 arrives by stop PI267 thanks
Out of order	"Plates", lights, in-transit, bus stop, supervise, off, signboard.	@Transantiago many buses have passed by with a sign saying IN TRANSIT in front of stop pf649 in Puente Alto
App	Bus stop, happens, information, metbus, service, redbus_urbano, busesvulesa.	@transantiago There is no information for service 110 at stop PI237. What's going on?
Social	Http, home, region, Metropolitan, Santiago, swarmapp.com, sq.com, fb.me.	Good morning ... it is so cold. Brrrr ... (@Stop PI242) https://www.swarmapp.com/paalejgonzalezm /checkin/53832ac4498e151d81ed4b65? s=2xbsdqx0yJopwJXTBMWNHGtStU&ref= tw%20â€¦

RQ2: Can tweets referring to the system be grouped into topics or themes?

As it is common with topic models, there are several topics that express the same concept with slightly different words. Thus, the generated topics can be fused into more semantically meaningful "thematic" or "aggregated" topics. In order to assign a document to its corresponding thematic topic, the probabilities of belonging to analogous topics were added and this sum was defined as the probability of belonging to the thematic topics. Tables 1 and 2 present the final aggregated topics and the best translation of the Spanish keywords in descending order of importance for each model. The final model of the Stops database has 10 topics, while the one of the Buses database has a total of six.

For both databases, the keywords are clear and are related to their respective topic. Particularly, the word "Plates" in the topics Stop and Out of order in Tables 1 and 2, refers to the different license plates of buses. For example, "cjrj", "bjfd", "cjrg", among others. In all cases, keywords were reduced to their semantic root. Although the different variations of the words could have been reduced to a common root prior to the topic modeling process, by performing a lemmatization step, it was ruled out since the variations of the words were minimal.

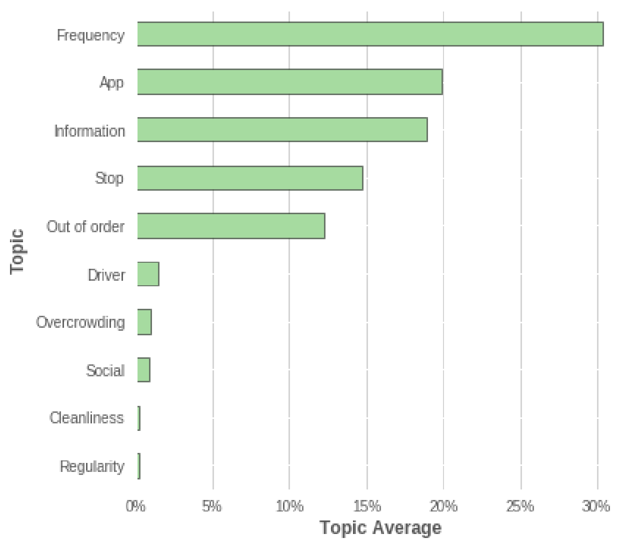


FIGURE 2. Topic distribution, Stops database.

Figures 2 and 3 show the topics distribution for the Stops and Buses databases, respectively. In Figure 2, the preponderant topics are not surprising: Frequency, Information, Stop, Out of order. Regarding validation, we found that the proportion of tweets and topics related to the real-time information app showed similar values, 18.4% for tweets and 19.9% in the topic model, largely validating the model.

The preponderant topics obtained by the model trained with tweets coming from the Buses database (Figure 3) are similar to the Stops database topic model, with the exception of the Stop topic, which shows greater importance. In terms of the accuracy of the model, the proportion of tweets related to

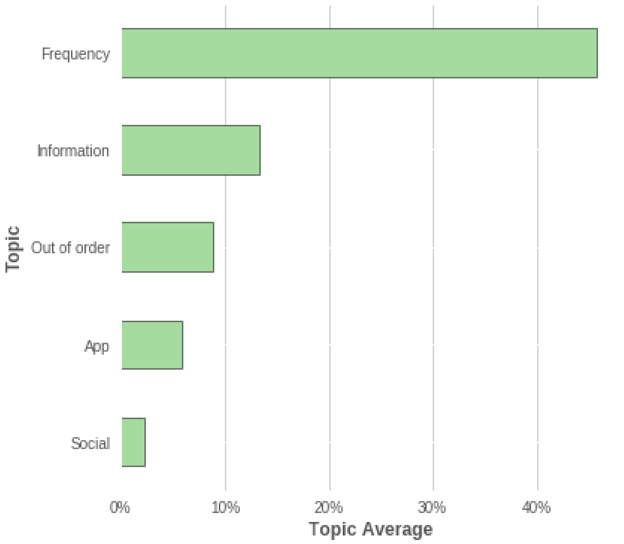


FIGURE 3. Topic distribution, Buses database.

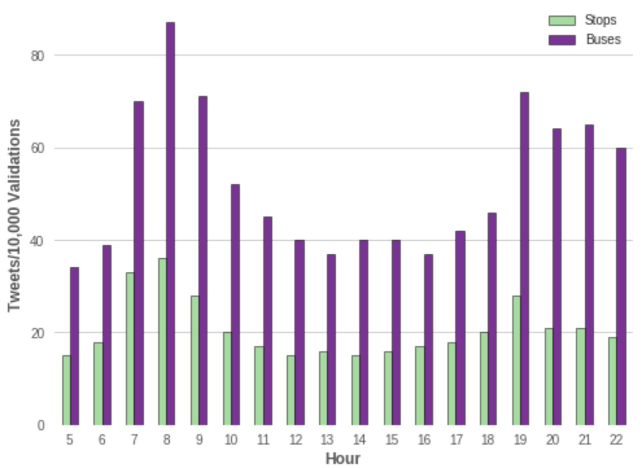


FIGURE 4. Temporal distribution tweets/10,000 validations, work day.

the information app is 7.4%, and the probability of occurrence of the App topic is 5.9%. Again, we consider this difference as minor.

RQ3: Is there a bias in using Twitter to measure the satisfaction of Transantiago bus users?

We performed a validation analysis based on three different aspects, namely temporal, spatial and operational.

Figure 4 shows the hourly evolution during the day of the number of tweets per 10,000 validations computed for both databases. A validation occurs when a smartcard is used to get on a bus. That is, is an indicator of the actual demand for buses, without considering evasion.

Both databases present a similar behavior throughout the day. At morning rush hours (between 7:00 and 10:00) and evening rush hours (between 18:00 and 21:00) there is a greater number of tweets per validation compared to other times during the day. Therefore, assuming the number of transfers per validation remains relatively constant during the day, the user of morning and evening rush hours is more active in the use of Twitter. Given the answer to RQ1, the above could also be explained by a worse performance of the system in those periods.

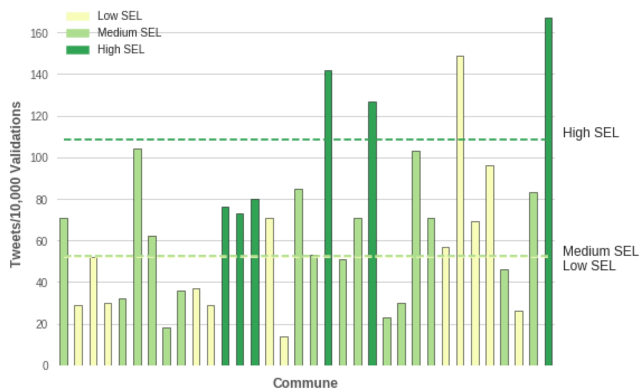


FIGURE 5. Tweets/10,000 validations per commune, 6 to 9 AM.

The graph in Figure 5 shows the same ratio as in Figure 4 but aggregated per commune, differentiated according to their socioeconomic level (SEL). The analysis was limited to the time slot between six and nine AM to study the behavior of the residents of each commune. Although the tweets and validations in this time slot can include users who transfer from other communes, it is assumed that in those communes the transfers at those hours are less than the validations of the trips originated there.

The average ratio of each socioeconomic level (calculated as the division between the sum of tweets per commune and the sum of the validations per commune belonging to each SEL) is shown in the horizontal lines in Figure 5. Trips starting in communes with a high SEL exhibit a greater tendency to use Twitter in comparison to trips starting in other communes. This suggests that the analysis could be biased towards the sectors of high socio-economic level of the city.

The final operational analysis compares the Frequency Compliance Index (*ICF* in Spanish) calculated by *DTPM* with the results of the model with respect to the *Frequency* topic. The *ICF* compares the effective versus the scheduled number of buses for each service for a given period. That is, it is a measure of bus frequency fulfillment. Each point in Figure 6 represents a bus operating firm in Transantiago.

The *ICF* (vertical axis) corresponds to the monthly average obtained between 2014 and 2016 (the higher the value, the better). The horizontal axis presents the probability of a tweet referring to a service of the corresponding firm to belong to the *Frequency* topic according to the model. Figure 6 shows that the *Frequency* topic is more relevant (higher percentage assigned by the model) in those firms with lower *ICF*. Therefore, the results of the proposed methodology are consistent with one of the performance indexes used by the authority to evaluate the operating firms.

RQ4: Can satisfaction surveys be replaced by information reported on Twitter?

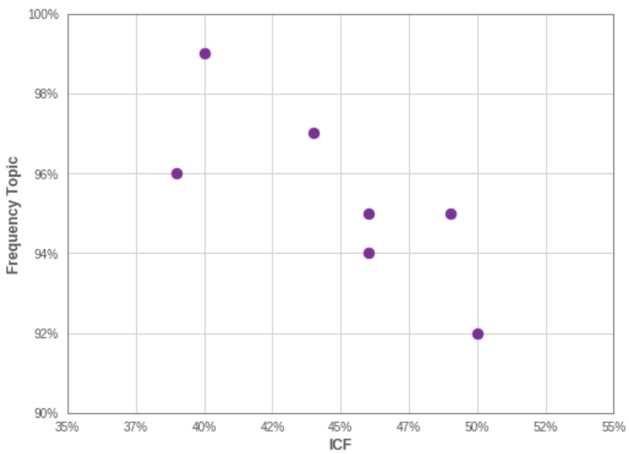


FIGURE 6. ICF operational analysis.

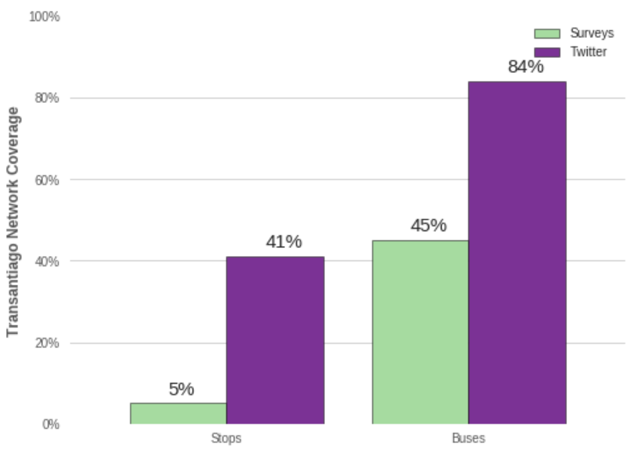


FIGURE 7. Transantiago network coverage.

The first aspect compared between both approaches was its network coverage, in terms of stops and bus services with information available. By 2016, the network had 11,340 bus stops and 379 bus services. As expected, Figure 7 shows that Twitter achieves a much greater spatial coverage than satisfaction surveys in both datasets.

During the three-year period analyzed, the surveys covered (by design) a total of 611 stops and 170 bus services, while the proposed technological tool has information for 4,640 stops and 318 bus services. However, the average number of tweets per stop is much lower than the number of surveys (six tweets versus 45 surveys). As compared to traditional survey, the proposed tool provides a higher coverage but a smaller amount of data per stop. For the bus services, however, the average amount of tweets per service is higher than the number of surveys per service (202 tweets versus 162 surveys).

Figure 8 compares the number of different people who contributed with information through both tools. It was assumed that satisfaction surveys did not repeat respondents, so one survey is equivalent to one respondent. On average, each

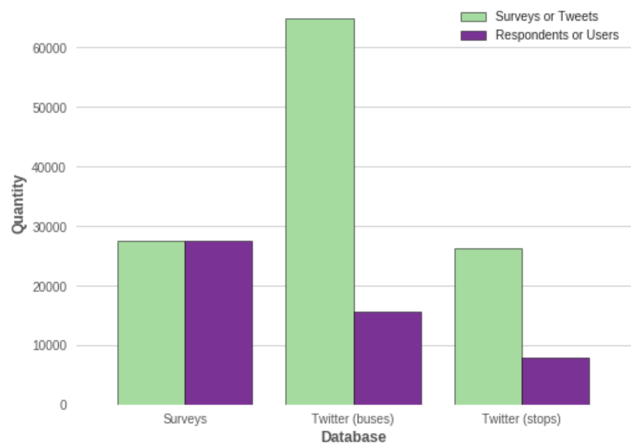


FIGURE 8. Respondents and users.

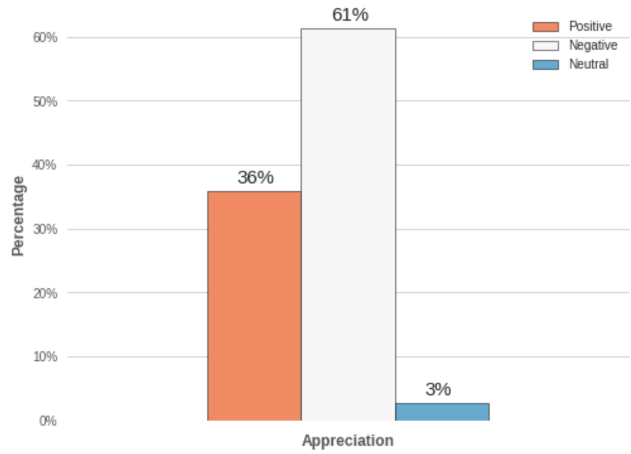


FIGURE 9. Sentiment analysis of surveys.

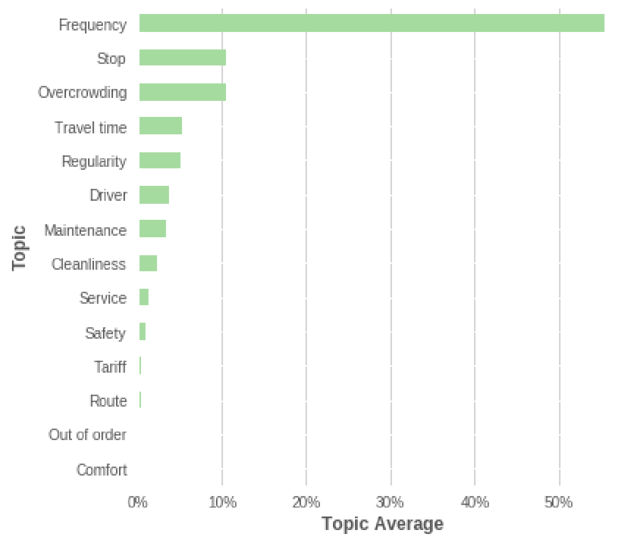


FIGURE 10. Topic distribution, surveys database.

Twitter user *tweets* 3 to 4 times during the three-year period analyzed. Satisfaction surveys reach a greater number of people than either of the two tweets databases. Therefore,

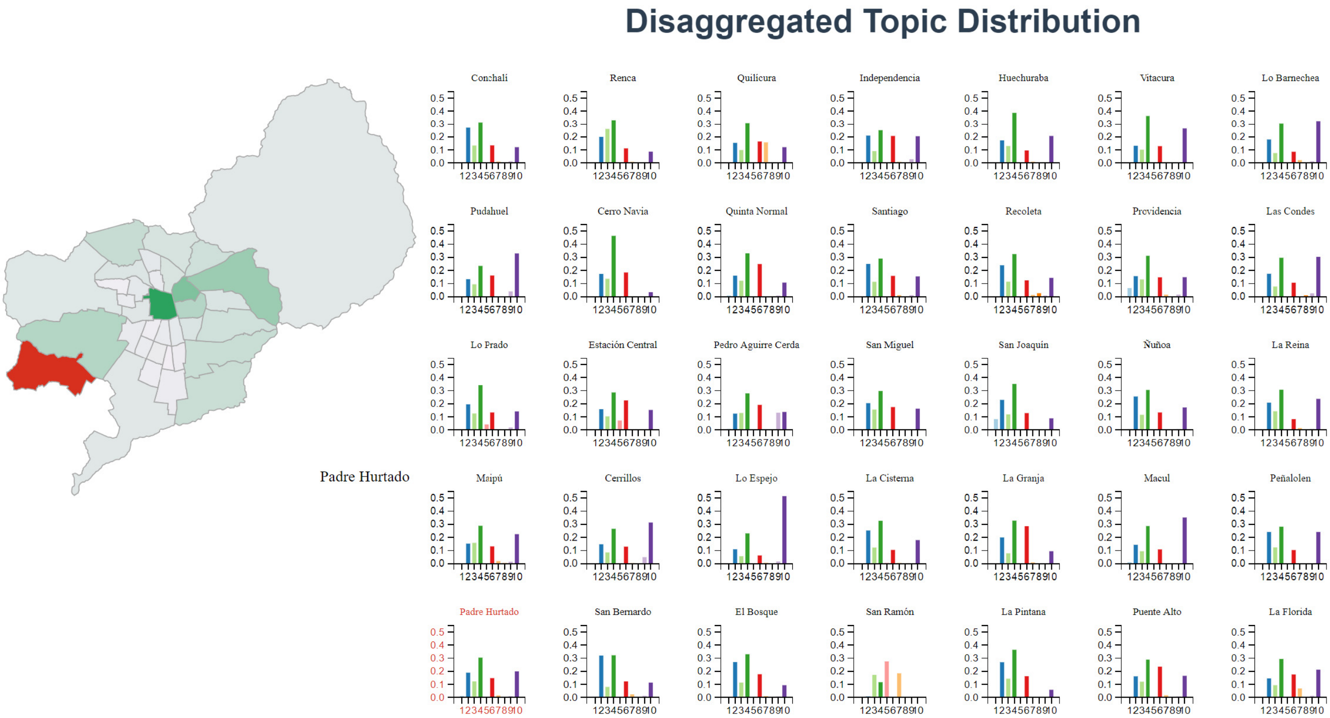


FIGURE 11. **Topic distribution by commune, Stops database.**

surveys can capture the opinion of the respondent at that time instant, while Twitter collects different manifestations over time from a smaller number of users.

Satisfaction surveys request respondents to qualify the service received in their travel experience and specify the reasons behind the assigned grade, which can be positive or negative. Figure 9 shows the distribution of the positive and negative reasons given by the respondents. The sentiment distribution in Figure 9 has a much greater presence of positive reasons than those of Twitter (Figure 1), in which it was less than two percent. This suggests that the satisfaction survey, by randomly choosing their respondents, allows evaluating both the negative and positive aspects of the user's travel experience, unlike the proposed tool that is limited mainly to negative topics.

The reasons given by the survey respondents to justify their negative evaluation were manually classified by topic. It is worth noting that the topics were defined according to our judgement. Figure 10 shows the topic distribution of the reasons given by the respondents to negatively qualify their bus service. Note that only the 61.2% of the respondent that negatively assess the system are considered here. Even though frequency is the main topic mentioned in the surveys, the topic distribution in Figure 10 and those from the topic models (Figures 2 and 3) are not similar. New topics arise which were not identified in the topic models. Thus, satisfaction surveys present a greater variety of reasons with which to justify the user's dissatisfaction with the service.

As a final experiment, we performed an analysis of the predominant topics in each commune and type of bus, for both tweets database and for the satisfaction surveys. While surveys tend to report average values for most of the topics in communes and operating firms, according to our results, tweets are able to capture certain different and valuable patterns. As an example, tweets show sectors of the city (West, North and South) where there is evidence of high user dissatisfaction. Moreover, certain firms obtain a high dissatisfaction, while others have good levels of satisfaction, something that is not easily observed by just analyzing surveys.

In order to further understand the topic structure, we illustrate the topic distributions in each commune of Santiago in Figure 11. We include a map of Santiago disaggregated by commune and their respective topic distributions. Communes in darker green tones have the highest tweet concentration. The figure includes 35 topic distributions, one for each of the 34 communes analyzed and a dummy commune (highlighted in red on the map of Figure 11) with the average topic distribution for the Stops database. We also developed an interactive version of the map,[5] where users can easily compare the average distribution of the whole tweet database with the distribution of a particular commune. This analysis allows us to identify topics that separate themselves from the average, implying a higher user dissatisfaction regarding that topic in that commune.

[5] https://topicos-reclamos-2018.herokuapp.com/#/barchart, developed by María Fernanda Sepúlveda (mfsepulveda@uc.cl).

VI. CONCLUSIONS

At the beginning of the article, four research questions were raised and answered throughout this investigation.

RQ1: Is Twitter mostly used by Transantiago bus users to express their discontent with the system?

Yes. While satisfaction surveys capture both the negative and positive aspects of the user's travel experience, Twitter focuses mainly on negative topics.

RQ2: Can tweets referring to the system be grouped into topics or themes?

Yes. In both databases analyzed, the key words are clear and are related to their respective topic. Additionally, the topic distributions are consistent with the reality of Transantiago.

RQ3: Is there a bias in using Twitter to measure the satisfaction of Transantiago bus users?

Based on the analysis of the presented variables (temporal, spatial and operational), it is possible to assert that the proposed methodology is valid to measure the dissatisfaction of Transantiago bus users. All the variables analyzed give reasonable results. However, two possible biases were identified, besides the one concluded in the first research question: i) users tend to tweet more on rush hours, and ii) communes with a higher socioeconomic level tweet more than those of lower incomes. These two issues must be taken into consideration when using the proposed tool.

RQ4: Can satisfaction surveys be replaced by information reported on Twitter?

Surveys and Twitter yield different but compatible information. For this reason, we believe that the use of topic models on tweets of Transantiago is limited to a complementary role to the satisfaction surveys.

VII. LIMITATIONS AND FUTURE WORK

As mentioned before, no lemmatization was applied to the databases. It would be interesting to analyze if the results of the topic models vary or remain similar when using it.

The analysis was limited to the tweets generated in working days. However, applying the proposed tool on weekends could help identify problems and trends that are not appreciated in the operation of the bus system from Monday to Friday.

The sentiments expressed on Twitter by Transantiago bus users may vary over the years, making it necessary to recheck the conclusion of the first research question. The manual sentiment analysis is tedious and consumes resources. There is a need for a sentiment analysis software in Spanish, adapted to the Chilean dialect, with a high level of certainty.

Future work regarding this investigation is related to applying different variations of topic models. The models could be generated using biterm topic models [21], which have been developed for short and informal texts and have shown good performance.

An interesting line of investigation is to apply semi-supervised topic models to the analyzed databases. SeededLDA [22] allows the user to provide the model with a list of key words of topics that are expected to be identified. The results of both types of topic models could be compared and see if the results improve by providing the algorithms with more information.

Additionally, our methodology could be eventually adapted to deal domains different from public transportation, such as debates, speeches, elections, sport matches and movies, where knowing people opinions and satisfaction in a timely manner is highly desirable.

Finally, future work could be related to analyze the evolution of these expressions over time, unlike what was done in this investigation where all the information was used statically. There is a methodology capable of doing this called dynamic topic models [23].

By solving the limitations of this research, exploring the proposed research lines and developing an easy-to-use software, what has been developed in this project has the potential to be used as a diagnostic tool complementary to the user satisfaction surveys currently carried out by the local transport authority.

ACKNOWLEDGEMENT

The Authors thank María Fernanda Sepúlveda (mfsepulveda@uc.cl) for her valuable help developing the topic visualization presented in Figure 11 and Hernán Valdivieso (hfvaldivieso@uc.cl) for his valuable help generating new visualizations and enhancing the figures presented.

REFERENCES

[1] *Directorio de Transporte Público Metropolitano*, Informe de Gestión, Santiago, Chile, 2016.

[2] *Encuesta de OpiniÃşn Pública: Evaluación Gestión de Gobierno*, Adimark, Santiago, Chile, 2017.

[3] C. Aggarwal, *Mining Text Data*. Springer-Verlag, 2012.

[4] *Estudio SatisfacciÃşn de Empresas Operadoras*, Brújula, Informe de Resultados, Santiago, Chile, 2016.

[5] M. Galesic and M. Bosnjak, "Effects of questionnaire length on participation and indicators of response quality in a web survey," *Public Opinion Quart.*, vol. 73, no. 2, pp. 349–360, 2009.

[6] P. C. Zegras, K. Butts, A. Cadena, and D. Palenci, "Spatiotemporal dynamics in public transport personal security perceptions: Digital evidence from Mexico City's periphery," in *ICT for Transport: Opportunities and Threats*. Cheltenham, U.K.: Edward Elgar Publishing, 2015.

[7] B. K. McHugh, B. Dong, J. D. Recker, and V. Shank, "Conducting onboard transit rider surveys with electronic handheld tablets: An agencywide consolidated approach," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2643, no. 1, pp. 19–27, 2017.

[8] New Cities Foundation. (2012). *Connected Commuting: Research and Analysis on the New Cities Foundation Task Force in San Jose*. [Online]. Available: http://www.newcitiesfoundation.org/wp-content/uploads/New-Cities-Foundation-Connected-Commuting-Full-Report.pdf

[9] K. Fárkas, Á. Z. Nagy, T. Tomás, and R. Szabó, "Participatory sensing based real-time public transport information service," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM WORKSHOPS)*, Mar. 2014, pp. 141–144.

[10] J. Arriagada, M. A. Munizaga, F. Hernández, A. Antoine, and M. Zúñiga, *Qué Tipo de Información se Puede Obtener de Una Aplicación móvil de Transporte Público? En XVIII Congreso Chileno de Ingeniería de Transporte*, La Serena, Chile, 2017.

[11] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[12] C. Collins, S. Hasan, and S. V. Ukkusuri, "A novel transit rider satisfaction metric: Rider sentiments measured from online social media data," *J. Public Transp.*, vol. 16, no. 2, pp. 21–45, 2013.

[13] D. Vilares, M. Thelwall, and M. A. Alonso, "The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets," *J. Inf. Sci.*, vol. 41, no. 6, pp. 799–813, 2015.

[14] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[15] L. Gao, Y. Yu, and W. Liang, "Public transit customer satisfaction dimensions discovery from online reviews," *Urban Rail Transit*, vol. 2, nos. 3–4, pp. 146–152, 2016.

[16] T. H. Chen, S. W. Thomas, and A. E. Hassan, "A survey on the use of topic models when mining software repositories," *Empirical Softw. Eng.*, vol. 21, no. 5, pp. 1843–1919, 2016.

[17] A. K. McCallum. (2002). *MALLET: A Machine Learning for Language Toolkit*. [Online]. Available: http://mallet.cs.umass.edu

[18] W. Zhao, J. J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, and W. Zou, "A heuristic approach to determine an appropriate number of topics in topic modeling," *BMC Bioinf.*, vol. 16, no. 13, p. S8, 2015.

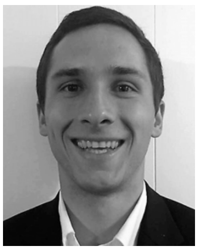
[19] P. F. Pai and C. H. Liu, "Predicting vehicle sales by sentiment analysis of Twitter data and stock market values," *IEEE Access*, vol. 6, pp. 57655–57662, 2018.

[20] F. Wang, K. Orton, P. Wagenseller, and K. Xu, "Towards understanding community interests with topic modeling," *IEEE Access*, vol. 6, pp. 24660–24668, 2018.

[21] X. Yan, J. Guo, Y. Lan, and X. Cheng, "A biterm topic model for short texts," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 1445–1456.

[22] J. Jagarlamudi, H. Daumé, III., and R. Udupa, "Incorporating lexical priors into topic models," in *Proc. 13th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2012, pp. 204–213.

[23] D. M. Blei, J. D. Lafferty, "Dynamic topic models," in *Proc. 23rd Int. Conf. Mach. Learn.*, pp. 113–120, 2006.



JOSÉ TOMÁS MÉNDEZ (M'18) received the B.S. degree in transportation engineering and the M.E. degree in transportation engineering from the Pontificia Universidad Católica de Chile, Santiago, Chile. His research interests include intelligent transportation systems, public transportation, user satisfaction, and topic modeling.

HANS LOBEL received the B.S. degree in computer engineering, the M.S. degree in computer science engineering, and the Ph.D. degree in computer science engineering from the Pontificia Universidad Católica de Chile, Santiago, Chile. Since 2016, he has been an Assistant Professor with the Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Santiago, Chile, and the Computer Science Department, Pontificia Universidad Católica de Chile. His research interests include visual recognition, machine learning, optimization, intelligent transportation systems, and big data.

DENIS PARRA received the B.S. degree in computer engineering from the Universidad Austral de Chile, Valdivia, Chile, and the Ph.D. degree in information technology and sciences from the University of Pittsburgh, Pennsylvania, USA. Since 2013, he has been an Assistant Professor with the Computer Science Department, Pontificia Universidad Católica de Chile, Santiago, Chile. His research interests include recommender systems, applications of data mining and machine learning, information retrieval, social network analysis, information visualization, and visual analytics.

JUAN CARLOS HERRERA received the B.S. degree in civil and industrial engineering from the Pontificia Universidad Católica de Chile (PUC), in 2002, and the MSc. and Ph.D. degrees in civil and environmental engineering (transportation program) from the University of California at Berkeley, Berkeley, in 2005 and 2009, respectively. He currently is an Associate Professor with the Department of Transport Engineering and Logistics, PUC. His research interests include traffic flow theory, traffic control, ITS, and traffic state estimation.