2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)
Bari, Italy. October 6-9, 2019

Listen to the People! Comparing Perceived and Documented Disruptions in Public Transportation, through Quantitative Quality of Experience, the Case Study of NYC

Gabriela Gongora Svartzman[1] and Jose E. Ramirez Marquez[2]

Abstract— **Public transportation systems are core infrastructures to smart cities and are expected to perform on a daily basis and in a reliable manner. However, such systems can be affected by external and internal events that manifest as delays. There are two sources customers can consult, which will impact their decision rewarding commuting times; either the official (provided by the transportation agency) or the reported by other passengers. It is debated whether the first approach is more accurate than the latter. This work compares both alternatives, using Natural Language Processing techniques and visual analytics to determine the delays on trains and evaluate their resilience process. For this purpose the impact of a massive snowstorm in November, 2018 to New York's Metro North Railroad system was used as a case study.**

I. Introduction

For the past decade public transportation agencies have been releasing data related to the performance of their services, i.e., schedules, routes, stations, number of passengers, etc. This data has made it easier to measure the quality of the services (e.g., how often the service experiences delays and for how long) through time ([1],[2]). In parallel, customers have reported on social media the quality of their experience riding the services, which at times includes delays, malfunctions, and service suspensions. Literature suggest ([3], [4], [5]) that customers' reports can be used to identify service disruptions and could highlight additional or different details than official sources. Therefore it becomes relevant to examine which of these data streams (or even their combination) can provide a more accurate and detailed account of the service, in particular during disruptive events to the service.

Work has been done to successfully identify and quantify disruptions in real-time, for similar city services and even entire cities when emergencies occur, such as fires, earthquakes and hurricanes ([6], [7], [8], [9], [10], [11]). Results have proven to be particularly successful when using Natural Language Processing (NLP) and topic analysis techniques, aiding in the categorization and estimation of disruptions to services ([4], [12], [13], [14]). Recent work suggests using these techniques to detect disruptions and traffic congestions in highways and roads ([2], [3], [15], [16]). The literature indicates a lack of research regarding the performance of public transportation systems and the comparison between traditional sensors that measure the quality of the service and the social media outlets that measure the quality of experience rewarding the service.

[1]Gabriela Gongora Svartzman is a PhD Candidate in Engineering Management, School of Systems and Enterprise, Stevens Institute of Technology, Hoboken, NJ, USA ggongora@stevens.edu

[2]Jose E. Ramirez Marquez is an Associate Professor and the Engineering Management Division Director, School of Systems and Enterprise, Stevens Institute of Technology, Hoboken, NJ, USA jmarquez@stevens.edu

This work proposes to bridge the quality of experience of customers with the quality of the services, in public transportation systems. Quality of experience being the qualitative measure of how the customers perceive the quality of the service, different to the quality of the service, which is the quantitative measure of how well the service is performing. The novelty of this research lies in the combination and contrast of the perceived and reported delays by customers in public transportation systems against the reported delays in real-time feeds by transportation agencies themselves. This comparison evaluates the resilience of transportation systems before, during and after disruptions. Differential commuting times are calculated, as well as the estimated waiting times per stage of resilience. Temporal visualizations are created to analyze and evaluate a snowstorm case study of New York's Metro North Railroad train system.

This paper is divided into six sections, where section I is the introduction to the research and section II describes previous work done, from traditional methods for traffic detection to event detection through social media. Section III provides context, describing the case study and data collected. Section IV then describes the methodology followed to bridge the delays reported on public transportation and the delays reported by the agency providing the service. All this leads to section V, which provides results on a specific example - a snowstorm in New York City - to illustrate the uses of the proposed methodology. At the end, section VI elaborates on discussions about the results and future work to be done.

II. Previous Work Done

A. Social Media for Crisis Management

In recent years there has been increased interest on the use of social media as a tool to help in crisis management, such as earthquake and hurricane relief. For example, Yates and Paquette [7] explore the use of social media for public participation in disaster response during the Haitian earthquake in 2010, emphasizing the need to broaden the use and scope of social media. Abel et al. [9] focus on crisis management through social media for identifying and fighting fire incidents. Their approach combines data from

Twitter and online news sources to verify incidents and report them.

Similarly, Schulz and Paulheim [17] use “mashups”, i.e., different streams of web and social media data to aid in crisis management (earthquakes, flooding, fires, among others). Considering information overload during a crisis, their mashup identifies the most time pressing issues, ranking them and filtering them, to enable faster decision making from staff members that are providing the help needed at the time needed.

B. Transportation services

Traditional methods for traffic detection and delay estimates in cities require the collection of sensor data to calculate the speed of cars around them, to determine traffic congestions [3]. Anantharam et al. [1] explore the possibility of combining social media with data streams from sensors to identify traffic related issues. The combined sensor observations with citizen’s reported observations allow an improved detection of traffic-related events and the authors urge to keep exploring this combination, as to gather further insights about the traffic-related events themselves.

In transportation services, Sasaki et al. [18] perform an initial exploration of Twitter as a train delay sensor, using the train lines in Tokyo. Twitter proves to be an accurate sensor and the authors suggest further exploration of Twitter data, overlaying additional data sources to strengthen the calculation and prediction of delays in trains. Seven years later, Ni et al. [16] focus on subway passenger flow prediction, through seasonal auto-regressive algorithms integrating moving average for time series prediction in turnstile data using event detection in social media. The authors prove that social media should be further exploited in traffic-related predictions.

C. Social Media and Disruptions in Transportation Services

Advances in text mining and Natural Language Processing (NLP) techniques have lead to applications of social media for event detection. Pereira et al. [19] use topic modeling and NLP techniques to extract key features from incident reports in real-time, subsequently feeding them into several prediction models to determine the clearance time between incidents. Overall, their models performed 28% more accurately when text analysis techniques are involved.

Gu et al. [15] detect incidents on highways and arterials through the use of Twitter. This analysis provides a more extensive coverage of the incident and additional incidents that were not recorded in the official 911 feed. Zhang et al. [20] also explore the importance of “Twitter concentrations” during traffic surges in North Virginia over a whole year. Their extraction, clustering and correlation algorithms uncover traffic patterns and traffic disruptions through social media, emphasizing on the importance of Twitter concentrations for their identification.

III. Data Description

Metro Railroad North is a system of trains operated by the Metropolitan Transportation Authority (MTA) in New York State. Different to the subway, the Metro North consists of trains that head north of New York, outside of New York City (NYC) and, into the suburbs. By the end of 2017 [21], the annual ridership for the Metro North Railroad was of over 86.6 million people, equivalent to an average of 298,300 weekday passengers, throughout 6 rail lines and 123 stations (or stops). This transportation system allows millions of people to commute into NYC and surrounding areas everyday, while living in the suburbs. These trains usually arrive every 10 to 60 minutes (depending on the rail and time of day), therefore a delay at any hour of the day can have a drastic impact in commuting times, different to the subway system where trains arrive with more frequency (2-10 minutes). Figure 1 shows an overview of the six Metro North Railroad Lines and their locations with respect to New York State and Connecticut.

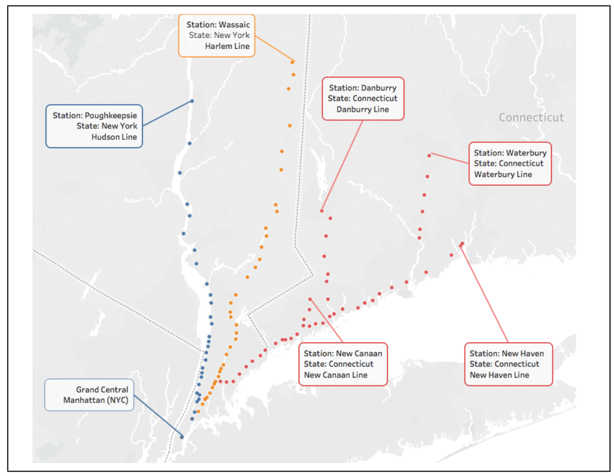

Fig. 1. Metro North Railroad Lines

This research obtains data from three different sources: General Transit Feed Specification (GTFS) files for train schedules, the official alert system for the Metropolitan Transportation Authority (MTA) service which controls the Metro North Railroad and social media (Twitter in this case). The first two data sources are official sources from the agency providing the service (the MTA), while Twitter relies on users to report incidents and provide complaints on the service.

The first stream of data collected are GTFS files, which record the schedule of the Metro North Railroad trains, i.e. when they arrived to a station, when they departed a station and so on. These files can be obtained from the MTA directly, by registering as a developer and following their file specifications. GTFS feeds can be static or real-time, the difference being that static feeds will provide the official schedule that the transportation service should follow, whereas the real-time feed reports the real schedule the transportation service followed. With this data it is possible to measure the delays in transportation systems.

The MTA also has a website where they post official alerts of the system, ranging from planned maintenance to unforeseen events such as a sick passenger, a snowstorm or a tree falling on the tracks. This feed is presented as a table with the date, the service affected (reports on all MTA owned services), the type of disruption and the description of the issue. The alerts sometimes continue throughout the resolution of the disruption, until the issue has been resolved. At this point the MTA issues an apology and the alerts cease until the next disruption on the service occurs. The MTA also provides an estimate of the delay, for example "...trains are running with 10 minutes of delay...". Nonetheless, customers on Twitter have claimed that alerts are at times inconsistent or out of timing with the service itself, thereby customers constantly check Twitter for a more accurate estimate of their commute.

IV. Methodology

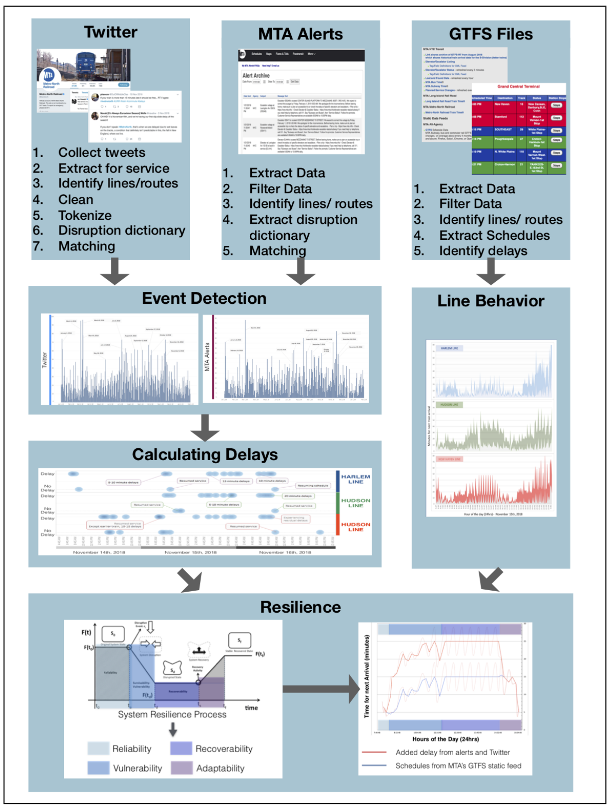

Fig. 2. Overview of methodology developed

The overall methodology followed is illustrated by Figure 2. In which the first part of the methodology evaluates the alerts and the tweets, i.e., the two data sources represented by text data. For the alerts feed, the following steps were implemented:

1) Identify the MTA alerts web feed, in which the data is presented in a four column table with the option of selecting a range of dates for the table.
2) The feed contained data in a dynamic context, therefore a dynamic web scrapper was required, with the following steps:
   a) Select dates from January 1st, 2018 to January 31st. These will be displayed as tables in several pages.
   b) At the end of the first page the total number of pages is displayed (e.g., 1 of 8567). Use this number to loop through the pages with all the tables.
   c) Obtain the data from the table.
   d) Go to next page, i.e., the next table.
   e) Repeat c) and d), adding rows to a single table (array), until the last page has been reached (number from step 2).
   f) Print entire table (assembled from all the tables) into a single file.
3) Sort the data (corresponding to the 13 months collected) according to the second column (service) and subset the rows that contain "MNR" in the second column. These are the alerts corresponding to the Metro North Railroad. The rest of the rows are alerts concerning the NYC subway, busses and other transportation services in New York, owned by the MTA.
4) Add another column to the subset from step 3, identifying the railroad line the alert corresponds to. This can be done using the third column of the table. For each row:
   a) Using Natural Language Processing (NLP) techniques [13] to tokenize the text, i.e., separate the text into words. If necessary transform to lowercase all tokens.
   b) Match each token against an array with all the names of the railroad lines. For example, if the tokens are ['Update', 'Hudson', 'line', 'experiencing', 'delays'] and they are compared against the array ['Hudson', 'Harlem',' Hudson',...], the tokens ['Hudson'..] and ['Hudson'...] will match, therefore this alerts belongs to the Hudson line.
5) The fourth and final column of the table contains a longer description of the disruption. In a separate array copy all the rows from this fourth column and convert them into a dictionary of disruptions by:
   a) Tokenizing every single rows and joining them together, similar to having a very long list of words or a Bag of Words [13].
   b) Removing stop words from a), i.e., words that join sentences and represent no information value. Context stop words are also considered (e.g., the names of the train railroads).
   c) Calculating the frequency of each token within a) and sorting the tokens in descending order. The words with highest frequencies are related to incidents occurring on the tracks, the trains, or to the whole system. An example of these words are ['delays', 'mechanical', 'vicinity', 'tree', 'police'].
   d) Repeating a) to c) using bi-grams (couple of words instead of single words) helped refine the dictionary created. Examples of disruptions in this case are ['police activity', 'issues vicinity', 'medical assistance', 'weather related', 'weather

advisory'...].

To extract the data from Twitter, hashtags for the Metro North Railroad had to be identified. The hashtags "metronorth" and "metronorthtweets" had thousands of tweets associated to them. Additionally the usernames @MyTransitApp, @cc_mnr_hud and @metronorth reported official delays of the Metro North Railroad. The tweets were extracted using the API provided by Twitter [22] then cleaned, removing: urls, emoticons and pictures. Tweets have more limitations because of their character restriction and informal language ([14]), nonetheless the tweets underwent the same process as the alerts, where the railroad lines affected (in the tweet) were identified and so were the disruptions. In this case, the disruption dictionary created with the alerts was used as a starting base to match tweets to disruptions. Several tweets could not be associated with a disruption from the dictionary, and in these cases the tweets were re-examined separately, and new disruptions were identified, further on added to the disruption dictionary.

The last section of the methodology involves the visualization and evaluation of the system's resilience process undertaken by transportation systems during a disruption to the system. Figure 3 exemplifies the resilience process taken by a system, based on Hosseini et al. [23] and Barker et al. [24].After a disruption occurs the time for the next arriving train will change, i.e. the system will behave differently. During the first stage, reliability, the system is behaving "normal". For this section it was intended to use the GTFS static files of the Metro North and compare them to the GTFS real-time files through a discrete event simulation. Unfortunately, after examining both data sources they appear to be the same, which means that the MTA does not store GTFS real-time historical files, only static files. For this reason, the static feed will be considered as the "normal" behavior of the system (as if the disruption had never happened).

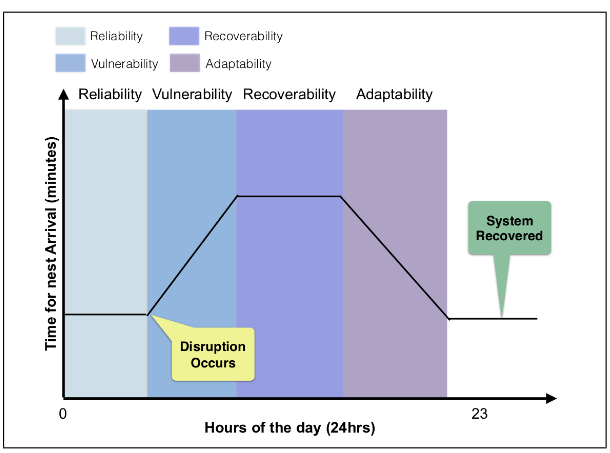

Fig. 3. Resilience Process and Measurement

To more accurately identify the effect of a disruption on the resilience of the system, the tweets and alerts were used. Through the disruption dictionary several events were identified. The beginning and ending of a disruption and the expected delay of the trains during a disruption can be obtained using the proposed methodology. Refer back to Figure 2 for overall methodology followed. A disruption was selected to demonstrate the proposed methodology, this being is the massive snowstorm that began on November 15th, 2018 affecting the East Coast (USA), in particular New York, New Jersey and Massachusetts. Only three railroad lines will be shown: New Haven, Hudson and Harlem. The next section will show the results obtained from this snowstorm.

V. RESULTS

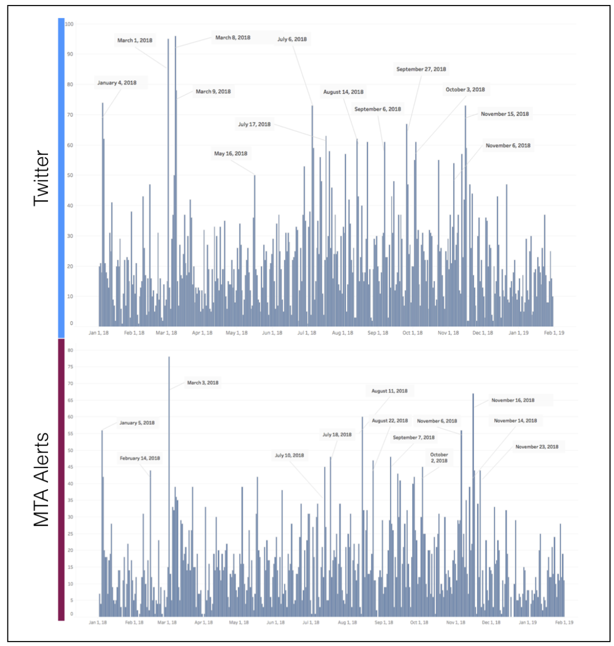

Fig. 4. Timelines of 13 months of data (January 2018 - January 2019) from Twitter and MTA alerts reporting incidents in any of the Metro North trains

Figure 4 shows a timeline of the number of records collected from Twitter and the MTA alerts reporting everything related to the Metro North trains. Both timelines follow a similar trend, although Twitter contains more reports and more frequent peaks than the MTA alerts. The majority of the tweets collected are from official sources, thereby they follow a certain resemblance to the MTA Alerts. Additionally, in twitter several commuters complain to official feeds about trains being delayed or other issues associated with the Metro North Railroad. The MTA then responds to customers about the status of the train or the particular issue from the complaint, therefore tweets can be a more accurate source in identifying the beginning of a disruption to the system, its process (resilience process) and the end of the disruption.

To exemplify how the the resilience process is determined from the combination of tweets and alerts Figure 5 shows the follow-up of delays (identifying the token "delay" in the text) throughout November 14th, 2018 up to November 16th, 2018. To make the visualization easier to understand only the alerts were plotted, identifying the beginning of disruptions, the size of the delays (time in minutes) and the end of disruptions. This visualization shows densities of alerts, which means that the darker the bubbles the more frequent the use of “delay” in the alerts extracted.

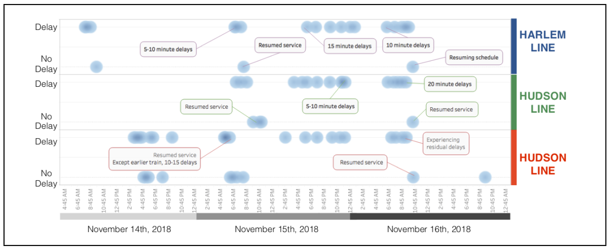

Fig. 5. Delays identified in twitter, throughout three different railroad lines. Dates: November 14th - 16th, 2018.

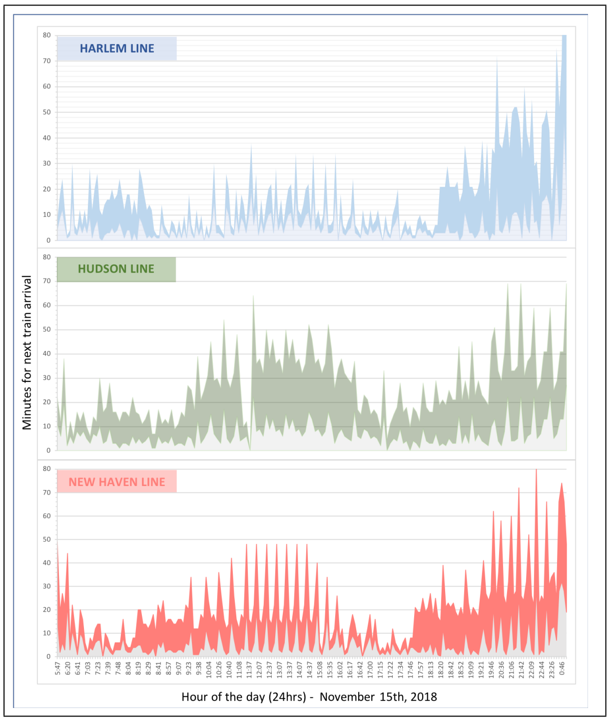

Fig. 6. Comparing Resilience in Metro North Railroad Lines during November 15th, 2018

Figure 6 shows the effect of the snowstorm on the New Haven, Hudson and Harlem, when combining all discussed data sources. To account for worst case scenario and be able to examine the results, this case study (and therefore Figure 6) only takes into account trains leaving Grand Central Station, as it is the largest station in the Metro North Railroad system and the place where all three lines start and end. Figure 6 shows the effects on Grand Central for each lines. The darker color on each of the three graphs represents the difference between the “normal” behavior (the ideal schedule of the trains, obtained from the GTFS static feed) and the delay on the trains reported by the combination of tweets and alerts. For the Hudson and New Haven lines, the discrepancy between both lines becomes evident during the middle of the day (8AM - 4PM). The more critical delays in all three lines are during the evening commute - 5:30 PM and onwards - as the delays go up to 80 minutes in some cases. This is the sum of time a customer would have waited for a train plus a delay originated from the disruption (i.e. the snowstorm in this case).

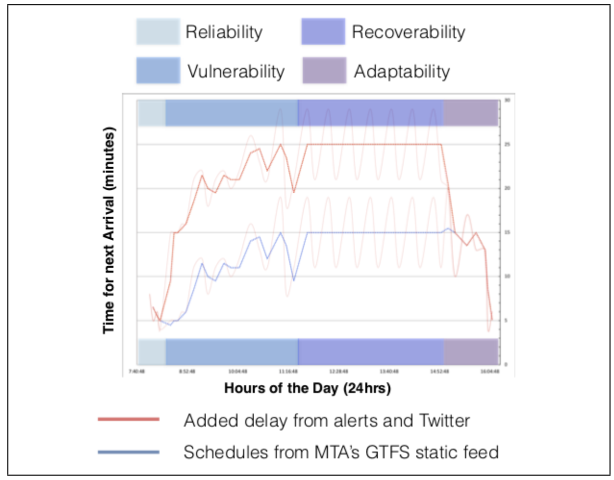

Fig. 7. Zooming into New Haven Line - November 15th, 2018. From 8AM to 4:30 PM.

TABLE I

RESILIENCE STAGES GRAND CENTRAL STATION, NEW HAVEN LINE ON NOVEMBER 15TH, 2018 (ZOOM-IN OF 8HRS OUT OF THE DAY)

Stages	Time of the day (hrs)	Waiting time (mins)
Reliability	before 8:14	6
Disruptive event	after 8:14	12
Vulnerability	11:44	25
Recoverability	14:55	16
Adaptabiliy	15:32	9

From Figure 6 it is evident that the system has a cascading effect from the snowstorm as it undergoes a few attempts at springing back into a reliable state. Zooming into one of this attempts is Figure 7, which visualizes the New Haven Line from 8AM to 4:30 PM. The red line is the estimated delays to the trains from the alerts and tweets. Moving average lines have been added to make the comparison easier (visually). Table I shows the exact times of each phase of the resilience process visualized in Figure 7.

VI. DISCUSSIONS AND CONCLUSIONS

This work consolidates a resilience analysis of train delays with the disruption detection of such event through social media and MTA alerts data. It is a first step towards a framework that can use multiple data sources to analyze a disruption from a qualitative and quantitative perspective, creating layers of detail and visualizing them to evaluate the resilience processes followed.

The main contributions of this research are the use of several data sources, joining them together and creating visual analytics from them. The use of different streams of data to measure and complement the same event and service, where each data stream adds an insight or angle a single stream of data would not satisfy by itself. Additionally, the development of visualizations and visual analytics has a data discovery role, as well as a tool to understand a complex system under real situations, in a cost-efficient environment.

This work exemplified the proposed methodology with one example but another contribution lies in the fact that the same framework is already built with thirteen months of data, therefore it is possible to zoom into any other disruption in that period of time, creating multiple points of interest and a variety of disruptions to analyze. The established methodology is the base for providing advisory on other routes to customers, or helping agencies detect flows on their transportation systems. Given a longer historical archive it will be possible to anticipate or predict the recovery time of railroad systems (and similar systems) based on different disruptions and their resilience analysis. With a prediction it will also be possible for the agency providing the service to deploy additional resources (e.g., busses to take customers to nearby stations) to improve their customer's experience while they fix or recover from a disruption.

Future work involves gathering more events, comparing different types of disruptions and their effects. Nonetheless, this point is subject to finding sufficient data during the same timeline for all data sources. Another aspect that should be explored, and it is suggested in the literature ([1], [4], [25]), is the use of other web streams. For example, the mining of online newspapers could be useful to automatically associate specific events to the disruptions occurring in the transportation system being measured. This point could even identify other services that might be suffering during the same disruption. Additionally data from weather sources becomes useful in places such as New York, where temperature can be extremely hot during the summer and extremely cold during the winter. These changes can account for a great percentage of the disruptions happening to the transportation systems.

ACKNOWLEDGMENT

The research performed by Gabriela Noemi Gongora Svartzman and Jose Emmanuel Ramirez Marquez leading to these results has received funding from the National Science Foundation, CRISP Type 2/Collaborative Research: Resilience Analytics: A Data-Driven Approach for Enhanced Interdependent Network Resilience, Award number 1541165.

REFERENCES

[1] P. Anantharam, K. Thirunarayan, S. Marupudi, A. P. Sheth, and T. Banerjee, "Understanding city traffic dynamics utilizing sensor and textual observations.," in *AAAI*, pp. 3793–3799, 2016.

[2] M. Liu, K. Fu, C.-T. Lu, G. Chen, and H. Wang, "A search and summary application for traffic events detection based on twitter data," in *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 549–552, ACM, 2014.

[3] T. Georgiou, A. E. Abbadi, X. Yan, and J. George, "Mining complaints for traffic-jam estimation: A social sensor application," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 330–335, ACM, 2015.

[4] K. Bontcheva and D. Rout, "Making sense of social media streams through semantics: a survey," *Semantic Web*, vol. 5, no. 5, pp. 373–403, 2014.

[5] A. Salas, P. Georgakis, and Y. Petalas, "Incident detection using data from social media," in *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on*, pp. 751–755, IEEE, 2017.

[6] K. Sasaki, S. Nagano, K. Ueno, and K. Cho, "Feasibility study on detection of transportation information exploiting twitter as a sensor," in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[7] D. Yates and S. Paquette, "Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake," in *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*, p. 42, American Society for Information Science, 2010.

[8] Z. Xiang and U. Gretzel, "Role of social media in online travel information search," *Tourism management*, vol. 31, no. 2, pp. 179–188, 2010.

[9] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, "Twitcident: fighting fire with information from social web streams," in *Proceedings of the 21st International Conference on World Wide Web*, pp. 305–308, ACM, 2012.

[10] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on twitter," 2012.

[11] M. Poblet, E. García-Cuesta, and P. Casanovas, "Crowdsourcing roles, methods and tools for data-intensive disaster management," *Information Systems Frontiers*, vol. 20, no. 6, pp. 1363–1379, 2018.

[12] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1079–1088, ACM, 2010.

[13] F. Atefeh and W. Khreich, "A survey of techniques for event detection in twitter," *Computational Intelligence*, vol. 31, no. 1, pp. 132–164, 2015.

[14] W. J. Corvey, S. Vieweg, T. Rood, and M. Palmer, "Twitter in mass emergency: what nlp techniques can contribute," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, pp. 23–24, Association for Computational Linguistics, 2010.

[15] Y. Gu, Z. S. Qian, and F. Chen, "From twitter to detector: Real-time traffic incident detection using social media data," *Transportation research part C: emerging technologies*, vol. 67, pp. 321–342, 2016.

[16] M. Ni, Q. He, and J. Gao, "Forecasting the subway passenger flow under event occurrences with social media," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1623–1632, 2017.

[17] A. Schulz and H. Paulheim, "Mashups for the emergency management domain," in *Semantic Mashups*, pp. 237–260, Springer, 2013.

[18] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*, pp. 851–860, ACM, 2010.

[19] F. C. Pereira, F. Rodrigues, and M. Ben-Akiva, "Text analysis in incident duration prediction," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 177–192, 2013.

[20] Z. Zhang, M. Ni, Q. He, J. Gao, J. Gou, and X. Li, "Exploratory study on correlation between twitter concentration and traffic surges," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2553, pp. 90–98, 2016.

[21] M. Metropolitan Transportation Authority, "The mta network: Public transportation for the new york region," *http://web.mta.info/mta/network.htm#statsmnr*.

[22] K. Makice, *Twitter API: Up and running: Learn how to build applications with the Twitter API.* " O'Reilly Media, Inc.", 2009.

[23] S. Hosseini, K. Barker, and J. E. Ramirez-Marquez, "A review of definitions and measures of system resilience," *Reliability Engineering & System Safety*, vol. 145, pp. 47–61, 2016.

[24] K. Barker, J. E. Ramirez-Marquez, and C. M. Rocco, "Resilience-based network component importance measures," *Reliability Engineering & System Safety*, vol. 117, pp. 89–97, 2013.

[25] M. Hu, S. Liu, F. Wei, Y. Wu, J. Stasko, and K.-L. Ma, "Breaking news on twitter," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2751–2754, ACM, 2012.