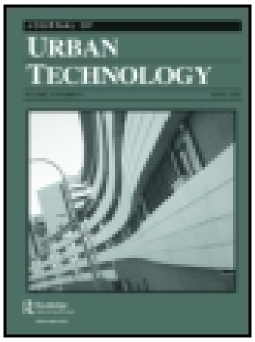
City Transit Rider Tweets: Understanding Sentiments and Politeness

Article *in* Journal of Urban Technology · February 2021

DOI: 10.1080/10630732.2021.1903288

Subasish Das
Texas State University

Hamsa Al-Zubaidi
University of Al-Qadisiyah-College of Engineering



Routledge
Taylor & Francis Group

Journal of Urban Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjut20

City Transit Rider Tweets: Understanding Sentiments and Politeness

Subasish Das & Hamsa Abbas Zubaidi

To cite this article: Subasish Das & Hamsa Abbas Zubaidi (2021): City Transit Rider Tweets: Understanding Sentiments and Politeness, Journal of Urban Technology, DOI: 10.1080/10630732.2021.1903288

To link to this article: https://doi.org/10.1080/10630732.2021.1903288

Published online: 26 Apr 2021.

JOURNAL OF URBAN TECHNOLOGY
https://doi.org/10.1080/10630732.2021.1903288

City Transit Rider Tweets: Understanding Sentiments and Politeness

Subasish Das [a] and Hamsa Abbas Zubaidi [b]

[a]Roadway Safety Division, Texas A&M Transportation Institute, Bryan, TX, USA; [b]Civil and Construction Engineering Department, Oregon State University, Corvallis, OR, USA

ABSTRACT
With the expanding popularity of Web 2.0, there has been a huge surge in the use of social media, like Twitter, to express user sentiments or opinions. Delays and breakdowns in transit operations can make riders annoyed and irritated, and as a result, they express their anger and frustration via social media posts. Understanding the tipping points of public frustration will help in developing better solutions. This study aims to develop a framework by developing multilevel sentiment analysis and determine the emotion and politeness measures using transit-related tweets from New York (New York City) and California (San Francisco). The popular hashtags associated with the transit systems of New York and California were collected during 2019. The words associated with negative sentiments widely differ in these two states. Moderate levels of differences are seen in the politeness measures for these two states. Additionally, co-occurrence measures associated with negative emotions identified unique issues based on the demographics. This study demonstrates that Twitter provides a great opportunity to understand the public perception of transit, and the findings can help authorities design a more efficient transit system to improve user experience.

KEYWORDS
Social media; Twitter; transit; sentiment analysis; politeness measure

Introduction

The impact of understanding public sentiments and opinions is crucial for transit agencies in data-driven decision-making processes. People's opinions and concerns about transit-related issues express the magnitude of real problems, especially in metropolitan cities. The common issues are related to either public transit systems or the problems they face while riding their own vehicles. Social media produces real-time big textual data that includes attitudes, opinions, and sentiments in different situations and events. In recent years, the role of social media has had various impacts on the field of transportation, which can indeed contribute to the process of decision-making for agencies.

According to a recent study (Cottrill et al., 2017), public transit services in the United States receive the highest number of negative tweets compared to other public and

CONTACT Subasish Das, PhD subasishsn@gmail.com Associate Transportation Researcher, Texas A&M Transportation Institute, 1111 RELLIS Parkway, Room 4414, Bryan, TX 77807, USA.

private services. This shows the significant amount of involvement required from people to express their concerns regarding transit services on social media, especially on Twitter. On the other hand, it was found that social media savvy transit operators have a higher number of positive comments or feedback. In fact, agencies that use Twitter to engage in conversations with users about their concerns or experiences of new or ongoing services have been associated with statistically higher positive sentiments on social media (Bregman, 2016). Although some agencies have represented this feedback, there is still a need for more data-driven approaches that oppose the current isolated management practices and address incorporating social media into ongoing transport planning, management, and operational activities (Schweitzer, 2014). There is a need for a data-driven analysis in order to understand the public reaction patterns towards their daily experience with the quality and performance of transportation services.

This study aims to answer two key research questions (1) **RQ1:** Do sentiments and reactions differ based on geographic locations? (2) **RQ2:** How do transit riders react on Twitter in terms of politeness measures? Understanding the sentiment and politeness dynamics of transit system riders from different geographic locations can help gain experience and knowledge on the relevant needs. The findings of this study can support research for planning and operations of public transit. This study collected Twitter data containing transit-related texts from New York (New York City) and California (San Francisco) from March to July in 2019. This study investigated network dynamics conducted on sentiment analysis and determined politeness measures for an in-depth understanding of transit user opinions and levels of satisfaction.

Literature Review

To comprehend the perception behind unstructured textual contents and its association in solving problems, researchers have applied sentiment analysis and opinion mining in many different transportation sectors. The relevant studies discussed below include three major research areas: (1) conventional survey analysis, (2) content analysis, and (3) sentiment analysis. Table 1 lists the key information identified from the studies discussed in this section.

Conventional Survey Analysis

With the objective of gathering a wide range of transit-related study issues that can be used to motivate potential public transit studies, Agrawal (2015) compiled findings from 56 US public opinion polls regarding public transit perspectives. This study defined the overall trends of transportation emerging across different studies in public opinion. Findings show that most people consider that transit brings many benefits, such as congestion relief and accessibility. Using the American Customer Satisfaction Index (ACSI) template, Shen et al. (2016) evaluated customer fulfillment for metropolitan train transportation in China.

Manville and Levine (2018) showed that most people consider the benefits of transit, such as improving environmental outcomes, reducing congestion, and easing accessibility. Among transit planners across Canada, Masood and Idris (2018) conducted a survey to cover the gaps of different transit stop factors and how to prioritize these factors. The

Table 1. Studies on transit-related public perceptions

Studies	Data Source & Type	Analytics	Approach/Findings
Conventional Survey Analysis			
Agrawal (2015)	56 surveys of US residents	Survey analysis	Findings show that most people consider that transit brings many benefits such as congestion relief and accessibility.
Shen et al. (2016)	American Customer Satisfaction Index (ACSI) model	Structural Equation Modeling (SEM)	Evaluated customer fulfillment, using SEM, for the metropolitan train transportation in China.
Masood and Idris (2018)	Canadian Urban Transit Association (CUTA) members	Expert opinion survey	The survey suggests that land-use is the factor that determines the position and spacing of transit stop, while real-time data is the most critical design factor for increasing ridership.
Manville and Levine (2018)	Survey data (1200 voters)	Survey analysis	Findings show that most people feel that transit brings many benefits such as improving environmental outcomes, reducing congestion, and easing accessibility.
Sarker et al. (2019)	1,369 responses from Innsbruck and Copenhagen	Survey analysis	The results show that the contributing factors for information sharing are social norms and self-actualization weighted against endeavor expectation.
Content Analysis			
Evans-Cowley and Griffin (2012)	Analyzed 49,000 posts on Twitter and other social media to examine public engagement	Social media mining	Results show that micro-participation via social media is effective in participation with substantial technical, analytical, and communication hindrances in influencing decision making.
Pender et al. (2014)	Tweets related to transit network disruptions	Twitter mining, Content analysis	The study suggested that much improvement is needed before using social media as an information delivery tool.
Lee et al. (2014)	Origin destination and Twitter data Southern California Association of Governments (SCAG) region	Tobit model	Results show the usefulness of harvested large-scale mobility data from location-based social media streams.
Nik-Bakht and El-diraby (2016)	Twitter follower analysis and profile development	Information retrieval method	Examined Twitter discussions with other online or offline means of public involvement in infrastructure projects.
Cottrill et al. (2017)	Tweets associated with @GamesTravel2014	Twitter mining, exploratory data analysis	This study evaluated both the structure and intent of the @GamesTravel2014 social media strategy via interviews with involved parties.
Casas and Delmelle (2017)	Twitter data from bus rapid transit system (BRT) in Cali, Colombia	Twitter mining, Content analysis	Findings identified several concerns of the riders including safety, system's infrastructure, and passenger behaviors.
Sentiment Analysis			
Collins et al. (2013)	Tweets about rapid transit system of the Chicago Transit Authority (CTA)	Sentiment analysis	More negative sentiment than positive sentiment.
Schweitzer (2014)	Large sample of Twitter comments	Sentiment analysis	More negative sentiment than positive sentiment.
Wu and Idris (2018)	Tweets of transit customers	Sentiment analysis	Tweets were visualized given their locations for problem detection and identification.
Haghighi et al. (2018)	Tweets on transit performance	Sentiment analysis, topic modeling	The tweet-per-topic index, as a measure of sentiment analysis, gauges transit riders' feedback and explores the underlying reasons behind dissatisfaction.

(*Continued*)

Table 1. Continued

Studies	Data Source & Type	Analytics	Approach/Findings
El-Diraby et al. (2019)	Data from the Twitter account of TransLink (Vancouver transit agency	Sentiment analysis	More negative sentiment than positive sentiment. However, sentiment levels in days with disruption showed lower levels of negative sentiments.
Qi and Costin (2019)	Tweets posted in Miami-Dade County	Sentiment analysis	The findings show that user habits (patterns of user's social interactions in Twitter) have great influence on sentiment value of selected tweets.
Li and Liu (2019)	26,000 comments posted on the Dazhong-Dianping website (Shanghai, China)	Sentiment analysis	The findings show that people are more satisfied with traffic hubs than vehicles. Bus comments reveal the lowest sentiment value, whereas comments about airports reveal the highest sentiment value.
Kim et al. (2020)	Large sample of Twitter data	Deep learning, Sentiment analysis	Developed a deep learning framework to capture local context among neighboring words in texts and is simplified by summarizing parameters in traditional models using a kernel function.

Delphi method was applied for content validity factor selection, and the survey yielded a content validity index of 0.78. The most significant planning factor for improving riding was real-time information, but the study also proposed that land-use was the main variable in determining the location and spacing of bus stations. Sarker et al. (2019) conducted research that depends on the willingness to transfer transport data as a portion of the normal routine use of transportation applications. Based on information collection that included 1,369 individuals from Innsbruck and Copenhagen as different towns in magnitude and overall cultural confidence, the empirical analysis consisted of the estimation of a structural equation model (SEM).

Content Analysis

Evans-Cowley and Griffin (2012) analyzed 49,000 Twitter, Facebook, and other social media data to examine public engagement in the Austin Strategic Mobility Plan. The findings show that micro-participation via social media is effective. However, substantial technical, analytical, and communication barriers remain in influencing policy and decision making. Pender et al. (2014) explored the role of social media in overseeing unexpected passenger transport disruptions through a global exercise study and an analysis of released studies. The findings indicated that the real-time aspect of social media might decrease the interrupted supply for transport. Lee et al. (2016) used a lately established Santa Barbara University algorithm and Twitter information to attain origin-destination pairs in the Greater Los Angeles Metropolitan Area known as the Southern California Governments Association (SCAG) region.

Nik-Bakht and El-diraby (2016) analyzed Twitter followers of a Light Rail Transit (LRT) project to explore means of public involvement in infrastructure projects. To communicate and provide transportation-related data and respond to the demands for data, Cottrill et al. (2017) examined the @GamesTravel2014 Twitter scenario to evaluate how this social media system was used during the 2014 Commonwealth Games in Glasgow, Scotland. This study assessed both the purpose and framework of the @GamesTravel2014 cultural press policy through account-related tweet evaluations and meetings with stakeholders. Casas and Delmelle (2017) accompanied a structured content analysis of the submissions with a text mining technique to the Bus Rapid Transit System scenario research in Cali, Colombia. The findings identified three main debate topics: problems with the infrastructure of the system, safety concerns, and behavioral issues on the bus. Public opinion was obtained from a Bus Rapid Transit System on Twitter. Rather than depending solely on automatic data mining methods to examine Twitter messages, the researchers used a two-step method in combination with a traditional qualitative research design.

Sentiment Analysis

Schweitzer (2014) investigated how the media depiction of public transit facilities might influence the way constituents and investors were planning their future transportation assets. From a large sample of Twitter posts, this study analyzed personal press content about government transportation, realized that it reflects more conflicting government transportation opinions than the remarks of most other government facilities, and included more adverse information about transportation customers. To assess the fulfillment of metro drivers, Collins et al. (2017) conducted a feeling survey acknowledging the restrictions of general efficiency metrics trends and tried to gauge the opinions of metro drivers by using Twitter link measurements. Conclusions were derived from standardized common feelings, the total positive and negative feelings, and the total number of tweets gathered over a period.

Wu and Idris (2018) analyzed the effectiveness of using Twitter information for visualizing and evaluating the fulfillment of travel clients through information mining, sentiment analysis, semantic analysis, and GIS visualization. The information used in this research was obtained from Twitter, employing a distinctive query mix with keywords such as the organization title and the method selection followed by a query language and regions. Haghighi et al. (2018) proposed a structure using Twitter information to evaluate the perspective of transportation drivers on the performance of transportation operation. To gauge the reviews of rail drivers and examine the fundamental factors that cause discontent with the system, an analysis of sentiment was completed further based on the tweet-per-topic test.

It is common that some people use social media as a medium to express their anger, frustrations, and negative sentiments. As a result, the inclusion of these negative comments without proper weightage may sway the research findings. By collecting tweets posted in Miami-Dade County during 2017 and 2018, Qi and Costin (2019) showed that the social media use patterns of users have a great influence on the sentiment value of selected tweets. Li and Liu (2019) analyzed 26,000 comments (posted on the Dazhong-Dianping website) about different transportation modes such as buses, rail transits, railway stations, and airports in Shanghai. Different text mining tools were applied to understand the commonness and characteristics of the different classes. El-Diraby et al. (2019) investigated the conceptual (what issues/topics are on customers' minds) network assessment and triangulation of people (how people are interrelated) in addition to the assessment of feelings (how they think about these subjects) of social media relationships to reinforce greater understanding of customer views and fulfillment rates of service. Kim et al. (2020) collected ride-hailing service-relevant text data from Twitter, created a database, and developed a novel Deep Learning (DL) framework that processes and classifies sentences that will automatically categorize the texts uploaded by service users according to transportation service-specific criteria.

The literature review reveals that several studies explored the potentials of examining customer feedback, opinions, and sentiments about transit experience. However, none of these studies focused on the determination of the emotion and politeness measures from transit-related social media mining. As studies have shown (e.g., Qi and Costin, 2019), there is a high likelihood of people using social media as a form of expressing negative views; therefore, there is a need to examine not only binary sentiments (either positive or negation), but also multilevel emotional contexts. This study applied innovative and state-of-the-art text mining tools in this unexplored field of transit-related studies.

Methodology

Data Collection

With approximately 500 million daily tweets, Twitter provides real-time big textual content with a wide range of themes and topics. The user posts, known as "tweets," cannot exceed 280 characters. Therefore, it not only disseminates information but also reflects opinions or sentiments within limited texts in real-time. This study used the open-source R software package *twitteR* to collect relevant tweets (Gentry, 2019). This study used the Twitter developer platform by using Open Authorization (OAuth), an authentication process that allows applications or tools to deliver client functionality to a web service without yielding an end user's identifications to the client itself, authentication as OAuth is mandatory for all Twitter-related data collection. The package can extract information on several variables, as listed in Table 2.

This study developed a comprehensive list of transit-related terms associated with California and New York transit systems. After collecting tweets from a wide range of transit-related social media accounts for both states, a list of significant key terms was identified. The final key search terms for collecting California (mainly San Francisco) transit-related tweets include *SFBART*, *metrolosangeles*, *SFMTA*, *SFTRU*, *CATransit*, *CASubway*, *losangelesbus*, *losangelessubway*, and *CABus*. The key search terms used for collecting New York transit-related tweets include *NYCTBus*, *NYCTSubway*, *NYPDTransit*, *NYTransit*, *NYBus*, and *NYSubway*. The time period of data collection was between June 2018 to June 2019. The number of collected unique tweets in New York and

Table 2 . Information collected using "Twitter" package

Analytics	Definition
Tweet	User or handle post (limited to 280 characters)
Handle	Username or profile in Twitter
Impressions	Times people were shown a tweet in the timeline or search results
Likes	Count of people who liked a tweet
Retweets	Count of resharing a tweet
Replies	Count of replies to a tweet
Timestamp	Timestamp of the tweet
Hashtags	Hashtags in a tweet

California databases was 51,356 and 10,344, respectively. For example, top four tweets with the highest number of retweets (using New York City data) are illustrated in Figure 1.

Concepts of Sentiment Analysis

Sentiment analysis is the computational study of people's emotions or opinions, and it is a challenging problem that is increasingly being used for decision-making by organizations and individuals. The three levels of sentiment analysis include document level, sentence level, and aspect level. These three levels of granularity are organized from coarsest to finest, with the finer granularity tasks being studied less.

This study used the open-source R software package *sentimentr* to develop the sentiment scores (Rinker, 2019). The augmented dictionary method of *sentimentr* provides better results than a simple lookup dictionary approach that does not consider valence shifter words (words that modify the connotation of the polarized words and include negators and amplifiers). The brief overview of the theoretical concept presented below is mostly based on Rinker (2019).

This method first uses a conventional senti-lexicon to tag polarized words and assign value to the polarity of each document or sentence. The algorithm uses each paragraph ($p_i = \{s_1, s_2, \ldots, s_n\}$) and breaks them into element sentences ($s_i, j = \{w_1, w_2, \ldots, w_n\}$) where w is the words within sentences. Each sentence (s_j) is broken into an ordered bag of words (a group of words; a representation of textual data) with the words as an i, j, k notation as $w_{i,j,k}$. The pause words or comma words are denoted as cw.

The words in each of the sentences ($w_{i,j,k}$) are examined and compared to a conventional sentiment lexicon (a dictionary providing scores for positive and negative scores based on the sentiment of the word; for example, the word "good" is associated with a positive score). In most cases, positive and negative words are scored with a plus one or minus one score. The weight (z) values can be justified with amplifiers/de-amplifiers.

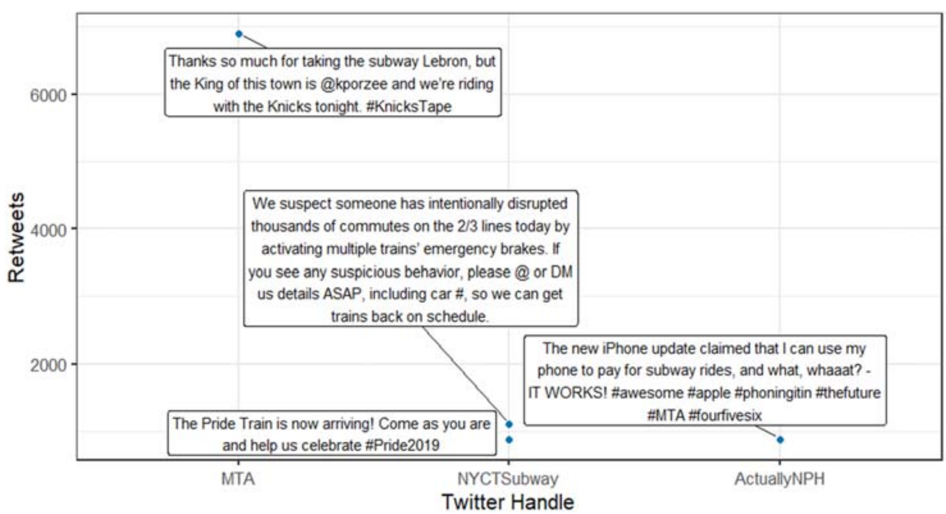


Figure 1. Top four tweets with the highest number of retweets from New York data

Rinker (2019) used the concept of cluster ($c_{i,j,l}$), a group of words to represent the contexts of polarity, which is a subset of a sentence. The overall goal is to determine an unbounded polarity or sentiment score ($\delta_{i,j}$), scores with no predetermined upper and lower limit thresholds, for each sentence. This score can be calculated by the ratio of summation of the clusters and the square root of the word count. For a comprehensive review of this concept, readers are referred to the study conducted by Rinker (2019).

Concepts of Politeness Measure

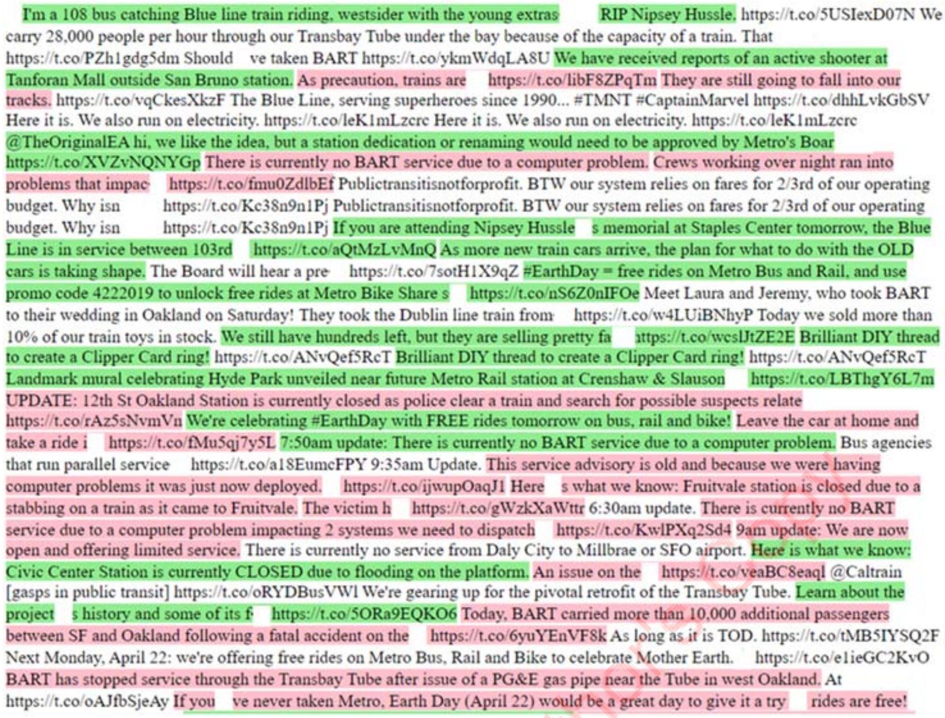
For the computational framework of politeness measures, the study employed the concepts developed by Danescu-Niculescu-Mizil et al. (2013). This study compared two classifiers: a linguistically informed classifier (LIC) and a bag of words classifier (BOW) and used human labelers as a reference point. The BOW classifier is a Support Vector Machine (SVM) using a unigram (single word) feature representation. The LIC classifier is an SVM using the linguistic features (for example, gratitude, deference, greeting, please, apologizing) in addition to the unigram features. This classifier shows higher levels of accuracy than the BOW classifier because of its input of additional features and contexts. For example, a sentence associated with features representing any kind of emotions can be better classified by an LIC classifier because of its use of linguistic contexts. Interested readers can consult Danescu-Niculescu-Mizil et al. (2013) for additional details of the classifiers.

Results and Discussions

Sentiment Analysis

As mentioned earlier, the research team used R software package *sentimentr* to develop the sentiment scores. This study developed a weblink to show the sentiment patterns of the tweets generated by the state (screenshots are shown in Figure 2). The green color indicates the tweets with some positive assertion. On the other hand, the red highlighting indicates the negative assertion in the associated tweets. For example, a tweet with three positive words, one neutral word, and five negative words will be assigned as a negative tweet because of the higher presence of negative words. Examples are detailed below:

- Line 9, Figure 2: "There is currently no BART service due to a computer problem"—this line contains a negative sentiment (highlighted in red) as two words are negative (no and problem), and one is positive (service)
- Line 21, Figure 2: "7:50 update: There is currently no BART service due to a computer problem"—this line contains a positive sentiment (highlighted in green) as two words are negative (no and problem) and three are positive (7:50 as real-time positive, update, and service). However, this real-time update is not a positive tweet. Based on the current sentiment lexicon, the assertion is real-time information sharing, which is positive since it will help commuters to plan accordingly based on the updated information. The current study used exiting sentiment lexicons to perform this task. There is a need for the development of a transit-based sentiment lexicon, which has not been developed yet, with weightage feature to minimize these misclassifications.



I'm a 108 bus catching Blue line train riding, westsider with the young extras RIP Nipsey Hussle. https://t.co/5USIexD07N We carry 28,000 people per hour through our Transbay Tube under the bay because of the capacity of a train. That https://t.co/PZh1gdg5dm Should ve taken BART https://t.co/ykmWdqLA8U We have received reports of an active shooter at Tanforan Mall outside San Bruno station. As precaution, trains are https://t.co/libF8ZPqTm They are still going to fall into our tracks. https://t.co/vqCkesXkzF The Blue Line, serving superheroes since 1990... #TMNT #CaptainMarvel https://t.co/dhhLvkGbSV Here it is. We also run on electricity. https://t.co/leK1mLzcrc Here it is. We also run on electricity. https://t.co/leK1mLzcrc @TheOriginalEA hi, we like the idea, but a station dedication or renaming would need to be approved by Metro's Boar https://t.co/XVZvNQNYGp There is currently no BART service due to a computer problem. Crews working over night ran into problems that impac https://t.co/fmu0ZdlbEf Publictransitisnotforprofit. BTW our system relies on fares for 2/3rd of our operating budget. Why isn https://t.co/Kc38n9n1Pj Publictransitisnotforprofit. BTW our system relies on fares for 2/3rd of our operating budget. Why isn https://t.co/Kc38n9n1Pj If you are attending Nipsey Hussle s memorial at Staples Center tomorrow, the Blue Line is in service between 103rd https://t.co/aQtMzLvMnQ As more new train cars arrive, the plan for what to do with the OLD cars is taking shape. The Board will hear a pre https://t.co/7sotH1X9qZ #EarthDay = free rides on Metro Bus and Rail, and use promo code 4222019 to unlock free rides at Metro Bike Share s https://t.co/nS6Z0nIFOe Meet Laura and Jeremy, who took BART to their wedding in Oakland on Saturday! They took the Dublin line train from https://t.co/w4LUiBNhyP Today we sold more than 10% of our train toys in stock. We still have hundreds left, but they are selling pretty fa https://t.co/wcslJtZE2E Brilliant DIY thread to create a Clipper Card ring! https://t.co/ANvQef5RcT Brilliant DIY thread to create a Clipper Card ring! https://t.co/ANvQef5RcT Landmark mural celebrating Hyde Park unveiled near future Metro Rail station at Crenshaw & Slauson https://t.co/LBThgY6L7m UPDATE: 12th St Oakland Station is currently closed as police clear a train and search for possible suspects relate https://t.co/rAz5sNvmVn We're celebrating #EarthDay with FREE rides tomorrow on bus, rail and bike! Leave the car at home and take a ride i https://t.co/fMu5qj7y5L 7:50am update: There is currently no BART service due to a computer problem. Bus agencies that run parallel service https://t.co/a18EumcFPY 9:35am Update. This service advisory is old and because we were having computer problems it was just now deployed. https://t.co/ijwupOaqJ1 Here s what we know: Fruitvale station is closed due to a stabbing on a train as it came to Fruitvale. The victim h https://t.co/gWzkXaWttr 6:30am update. There is currently no BART service due to a computer problem impacting 2 systems we need to dispatch https://t.co/KwlPXq2Sd4 9am update: We are now open and offering limited service. There is currently no service from Daly City to Millbrae or SFO airport. Here is what we know: Civic Center Station is currently CLOSED due to flooding on the platform. An issue on the https://t.co/veaBC8eaql @Caltrain [gasps in public transit] https://t.co/oRYDBusVWl We're gearing up for the pivotal retrofit of the Transbay Tube. Learn about the project s history and some of its f https://t.co/5ORa9EQKO6 Today, BART carried more than 10,000 additional passengers between SF and Oakland following a fatal accident on the https://t.co/6yuYEnVF8k As long as it is TOD. https://t.co/tMB5IYSQ2F Next Monday, April 22: we're offering free rides on Metro Bus, Rail and Bike to celebrate Mother Earth. https://t.co/e1ieGC2KvO BART has stopped service through the Transbay Tube after issue of a PG&E gas pipe near the Tube in west Oakland. At https://t.co/oAJfbSjeAy If you ve never taken Metro, Earth Day (April 22) would be a great day to give it a try rides are free!

Figure 2. Sentiment highlights by tweets (example from California data)

This unique visualization presentation provides a broad picture overview of the collected tweets. The average numbers of positive terms in both databases are positive (+0.091 for California and +0.082 for New York). These values answer research question one (*Do sentiments and reactions differ based on geographic locations?*) by providing evidence that politeness measures vary by cities.

The average profanity of New York's tweets is higher than California's tweets. Table 3 depicts the distributions of the terms associated with different emotions based on eight Plutchik (1991) categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Instead of conducting binary classification (positive and negative sentiment), Plutchik's multilevel sentiment or emotion classification can provide additional contexts for the analyzed tweets. The negated emotions are either the prefix of emotion-based keywords or their relative presence before or after an emotion-based keyword. For example, happiness contains the emotion "joy," and "unhappy" is a "joy-negated term." The negated terms have lower rate distribution of the emotions than the non-negated emotion-related terms. The emotions associated with anticipation (showing enthusiasm) and trust are higher in frequencies than other emotion types. Strong negative emotions (for example, anger, disgust) show lower frequencies than positive emotions (for example, joy). However, other forms of negations (for example, fear, sadness) exist in the collected tweets. The emotion score represents the percentage measures of emotion words in that emotion type. This multilevel classification shows that binary sentiment analysis is not sufficient in gaining knowledge from transit-related tweets. The statistics measures in Table 3 answer research question one (RQ1).

Table 3. Descriptive statistics of the emotion scores

	Count		Emotion Score		Standard Deviation of Emotion Score	
Emotions	NY	CA	NY	CA	NY	CA
anger	653	776	0.0333	0.0303	0.0053	0.0068
anger-negated	25	33	0.0041	0.0061	0.0002	0.0003
anticipation	3145	2343	0.0771	0.0553	0.0258	0.0204
anticipation-negated	65	91	0.0085	0.0118	0.0005	0.0008
disgust	459	411	0.0324	0.0266	0.0038	0.0036
disgust-negated	15	18	0.0027	0.0039	0.0001	0.0002
fear	1052	874	0.0330	0.0301	0.0086	0.0076
fear-negated	43	44	0.0085	0.0084	0.0004	0.0004
joy	1726	1305	0.0741	0.0523	0.0141	0.0114
joy-negated	32	45	0.0070	0.0086	0.0003	0.0004
sadness	1058	1064	0.0350	0.0344	0.0087	0.0093
sadness-negated	49	45	0.0085	0.0081	0.0004	0.0004
surprise	1429	746	0.0716	0.0357	0.0117	0.0065
surprise-negated	24	26	0.0061	0.0048	0.0002	0.0002
trust	2983	2749	0.0771	0.0567	0.0244	0.0239
trust-negated	57	88	0.0075	0.0081	0.0005	0.0008

Valence Shift Word Graphs

Dodds and Danforth (2010) introduced the concept of "Valence Shift Word Graph" in comparing sentiments in different document categories. This visualization provides the ranks of words by their descending absolute impact to the shift in mean valence between the two groups or categories, δ. Word *i*'s contribution depends on its shift in relative count, and its valence relative to the other group (Dodds and Danforth, 2010). To compare some text *n* in regard to a given text m, the valence difference can be defined as:

$$\delta(\mathrm{n},\ \mathrm{m}) = \mathrm{vs_n} - \mathrm{vs_m} \tag{1}$$

where, $\mathrm{vs_n}$ = valence shift in *n*, and $\mathrm{vs_m}$ = valence shift in m; the percentage contribution to this difference by word *i* can be expressed as

$$\Delta_{\mathrm{i}}(\mathrm{n},\ \mathrm{m}) = 100 \times \frac{(\mathrm{p_{i,n}} - \mathrm{p_{i.m}})(\mathrm{vs_i} - \mathrm{vs_m})}{\delta(\mathrm{n},\ \mathrm{m})} \tag{2}$$

where $\mathrm{p_{i,m}}$ and $\mathrm{p_{i,n}}$ are the fractional abundances of word *i* in texts *m* and *n*. The sum of $\Delta_{\mathrm{i}}(\mathrm{n},\ \mathrm{m})$ over all *i* gives *a hundred percentage postive or negative scores* depending on whether $\delta(\mathrm{n},\ \mathrm{m})$ is positive or negative. Figure 3 can be seen to have the following interpretations:

- Words on the right contribute to an increase in positive emotions in the corpus. A right yellow bar with a down arrow indicates less used negative emotion. A right purple bar with an up arrow indicates more used positive emotion.
- Words on the left contribute to a decrease in position emotions in the corpus. A left yellow bar with an up arrow indicates more used negative emotion. A left purple with a down arrow indicates less used positive emotion.

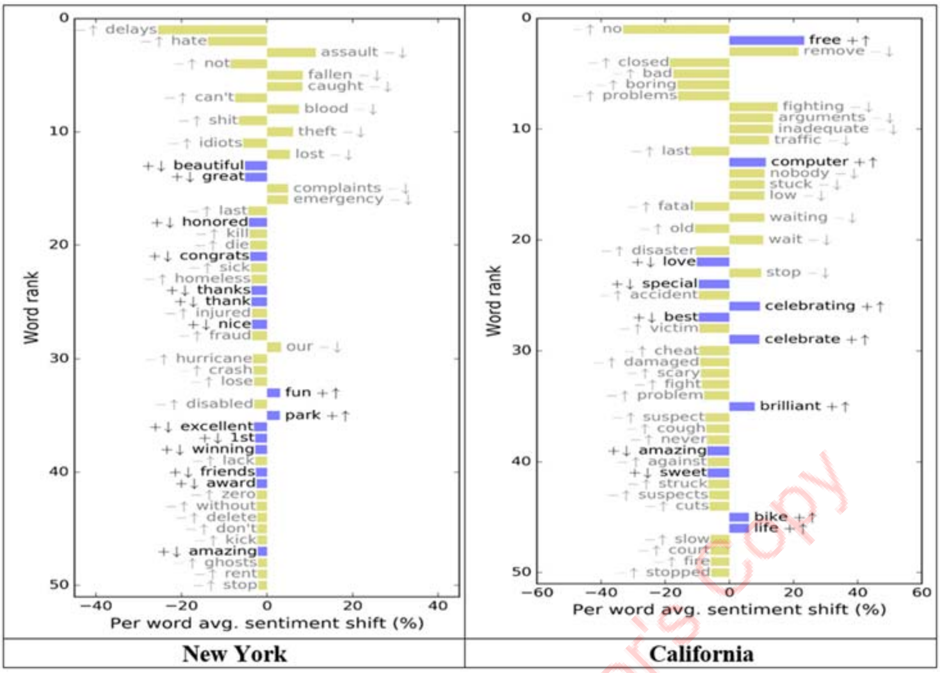


Figure 3. Valence shift word graphs

The plots show that positive sentiments are used more in California tweets compared to New York tweets. The positive sentiment associated words in California corpus are "life," "brilliant," "celebrate," and "free." In the New York corpus, two positive words ("fun" and "park") are used. Delays and hate are the two negative words with higher shift values in New York tweets. The shift values of the negative words less used are higher in California tweets than New York tweets. The findings from valence shift analysis help answer research question one (RQ1).

Co-Occurrence with Negative Words

It is crucial to understand the cause of the negative sentiments. Figure 4 shows the network plot of the words that are associated with negative terms in the "udpipe" sentiment dictionary (Wijffels, 2019*)*. According to the udpipe framework, the graph shows words of the dictionary in red and words that are linked to that word in another color. This is done by using the dependency relationship output to examine which words are linked to negative words from the "udpipe" dictionary (Wijffels, 2019). In New York, the term "sick" is represented with the darkest line in the figure followed by terms such as "dirty," "temporary," and "further." In California, the terms "homeless," "low," and "lose" are represented with the darkest lines in the figure. Moreover, some words that are linked to the words from the dictionary were "passenger," "customer," "car," and "maintenance" in New York's data. The California data contain words such as "income," "team," and "item." The relationship between "homelessness" and "transit-related issues" is noticeable. California is indeed one of the states with the highest population of homeless people, which is also increasing each year. There has been

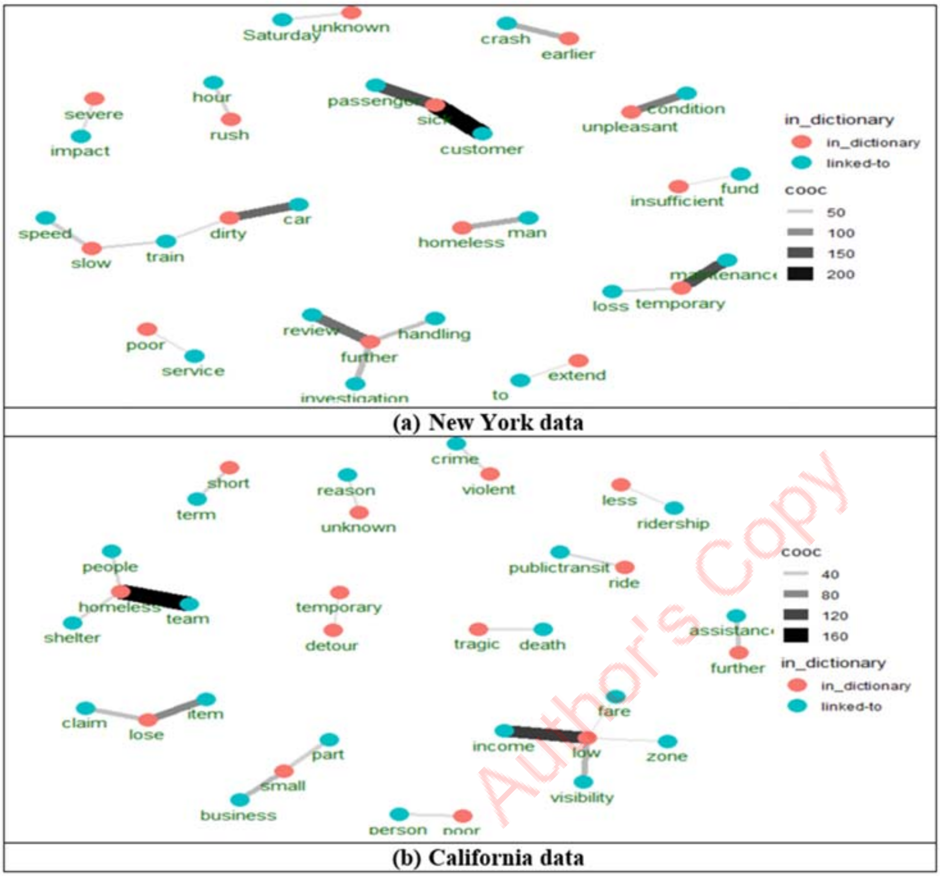


Figure 4. Network of words associated with negative sentiments

a growing concern regarding the constant presence of homeless people in public transportation like buses and trains. These concerns have been from both the users and from the drivers. People have expressed their concerns mostly in regard to their safety, and the drivers have expressed their concerns mostly about the homeless people who use buses and trains as their shelters. The L.A. County Department of Health Services has recently extended the Metro contract for two years for homeless outreach services. To address these issues, Metro launched a next-generation bus study to improve service. The co-occurrence plots show user reaction patterns to answer research question two (*How do transit riders react on Twitter in terms of politeness measures?*).

Measuring Politeness

Other studies showed that transit users express more negative sentiments than positive sentiments (Collins et al., 2012; Schweitzer, 2014). This study used an innovative approach by inspecting the results of the main politeness function to determine the percentage distribution of politeness among the documents, analyzing the documents associated with New York and California separately. Figure 5 illustrates how the

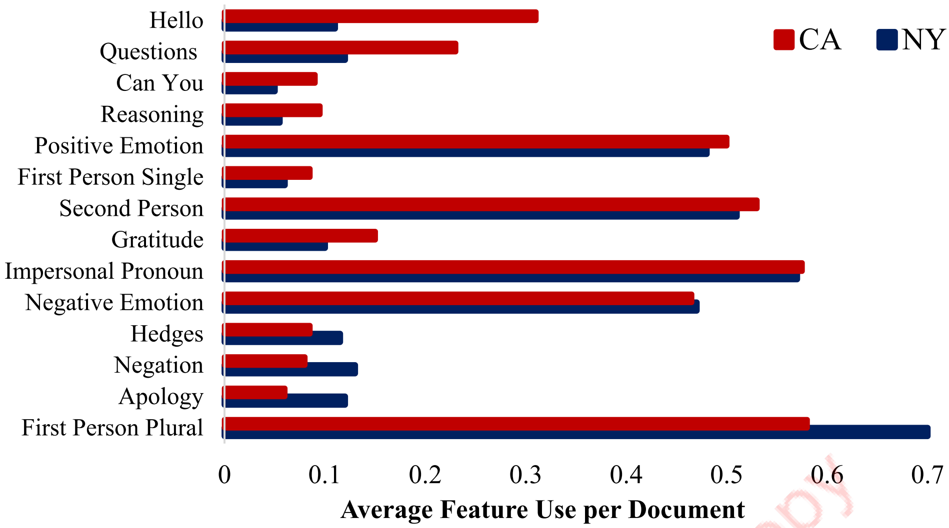


Figure 5. Politeness measures by state

frequencies of every politeness feature vary across a binary covariate of interest. The order of the features is determined by determining the variance-weighted log odds of each feature with respect to the binary covariate. Each feature is calculated using a t-test, and features are eliminated when the *p-value* of this test lies above the cut-off value employed by the users (Yeomans et al., 2019). A list of 36 different politeness features was introduced in Danescu-Niculescu-Mizil et al. (2013). This study found that 14 features are statistically significant out of the 36 different features. The politeness feature with the highest frequency in both states is "first person plural" (e.g., it will help us to go there quickly). Features such as "can you" (e.g., can you let us know earlier?) and "first person singular" (e.g., it will help me to go there quickly) were not frequently used in these documents. In most features, both states remained relatively close on average to one another. For politeness measures such as "first person plural" (e.g., it is a good deal for us), "apology," "negation," and "hedges" (e.g., I might use the blue line), New York had a higher average feature per document than California. For measures like "negative emotion," "impersonal pronoun" (e.g., service change reported on), "gratitude," "second person," and "positive emotion," the average feature per document was approximately similar for both states. California showed higher feature per document for measures such as "first person singular," "reasoning," "can you," "questions," and "hello." This section shows that linguistic difference of the politeness features varies by different locations. Research question one (RQ1) is addressed here because this analysis provides granular level of comparison between the politeness measures associated with these two cities.

The findings of this study provide the information needed to answer the research questions. The results show that binary sentiment analysis (positive and negative sentiments) is not always adequate in analyzing transit tweets as other emotions are also associated with these tweets. Two emotions (anticipation and trust) are dominant in frequencies in both datasets. The co-occurrence of negative words and politeness measures show a distinction between the two cities. The analysis shows the use of terms (for example, the word “homeless” in the California dataset and “dirty” and “unclean” in New York dataset) and the terms associated with politeness measures vary by different demographics. Findings from sentiment analysis, emotion mining, valence shift analysis, and politeness measure analysis show that the intersection patterns vary by state. The findings can help policy makers in determining the key issues from public sentiments and reaction patterns to resolve the issues in a quick fashion.

Conclusions

According to the recent statistics (Bregman, 2016), the first two significant transit-related concerns of US riders are the delays and safety of public transit. Social media are huge platforms for people to express their satisfaction, as well as dissatisfaction with regard to their experiences of using transit services. Understanding these concerns contribute to the ability of agencies to improve their services and make efficient decisions. Several studies have been conducted to analyze social media users’ opinions through polls, surveys, comments, and posts, which have been retrieved mostly from Twitter.

This study has selected an important topic that can help the academic community and practitioners to better understand customers and improve transit service. The key contribution of this study is that it developed a framework to extend binary sentiment analysis to a multilevel emotion analysis to analyze transit-related texts from New York and California riders. Two main research questions are answered by using different tools. For example, word co-occurrence analysis was performed to answer RQ2. To answer RQ1, several methods such as sentiment analysis, emotion mining, valence shift analysis, and politeness measure analysis were performed. Another contribution of this study is that it shows how interaction patterns vary by city at granular word level to understand the linguistic and interaction variation by cities and the transit services. This study applied eight key politeness measures to better understand transit user opinions, concerns, experiences, and levels of satisfaction. The results demonstrated that the politeness features and sentiments differ in both states for different contexts. The analysis of the co-occurrence of negative words can help authorities understand key issues, needs, and concerns. The methods used in this study are helpful for the use of social media-related knowledge in transit planning and operations.

The current study is not without limitations. First, the data collection period is limited, and it focused on only two states. A comprehensive analysis can be done using more years of data with the inclusion of more states. Second, the study is focused on the development of a text mining pipeline with the inclusion of innovative tools to answer the research questions. Each of the conceptual tools (for example, valence shift measure) can be explored more in depth to develop a stand-alone study. Future studies can use the current analytical pipeline to determine rider only emotions by removing tweets generated from the transit authorities’ official handles.

Acknowledgments

The authors would also like to thank two anonymous reviewers. We have tried to incorporate as many as possible of their insightful suggestions.

Notes on Contributors

Subasish Das, is an assistant research scientist with Texas A&M Transportation Institute in Bryan, Texas.

Hamsa Abbas Zubaidi, is a Ph.D. candidate in the Civil and Construction Engineering Department at Oregon State University, Corvallis, Oregon.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753

References

A. Agrawal, *What Do Americans Think about Public Transit? A Review of US Public Opinion Polling Survey Questions* (San Jose, CA: Mineta Transportation Institute, 2015).

S. Bregman, *TCRP Synthesis 99: Uses of Social Media in Public Transportation* (Washington DC: National Academies, 2016).

I. Casas and E. Delmelle, "Tweeting about Public Transit: Gleaning Public Perceptions from a Social Media Microblog," *Case Studies on Transport Policy* 5: 4 (2017) 634–642.

C. Collins, S. Hasan, and S. Ukkusuri, "A Novel Transit Riders' Satisfaction Metric: Riders' Sentiments Measured from Online Social Media Data," *Journal of Public Transportation* 16: 2 (2012) 21–45.

C. Cottrill, P. Gault, G. Yeboah, J. D. Nelson, J. Anable, and T. Budd, "Tweeting Transit: An Examination of Social Media Strategies for Transport Information Management during a Large Event," *Transportation Research Part C: Emerging Technologies* 77 (2017) 421–432.

C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts, "A Computational Approach to Politeness with Application to Social Factors," 2013 <https://nlp.stanford.edu/pubs/politeness.pdf>Accessed March 30, 2020.

P. Dodds and C. Danforth, "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents," *Journal of Happiness Studies* 11: 4 (2010b) 441–456.

T. El-Diraby, A. Shalaby, and M. Hosseini, "Linking Social, Semantic and Sentiment Analyses to Support Modeling Transit Customers' Satisfaction: Towards Formal Study of Opinion Dynamics," *Sustainable Cities and Society* 49 (2019) 101578.

J. Evans-Cowley and G. Griffin, "Micro-Participation with Social Media for Community Engagement in Transportation Planning," *Transportation Research Record* 2307, no. 1 (2012) 90–98.

J. Gentry, twitteR: R Based Twitter Client. R package version 1 <https://CRAN.R-project.org/package=twitteR>Accessed March 30, 2020.

N. Haghighi, C. Liu, R. Wei, W. Li, and H. Shao, "Using Twitter Data for Transit Performance Assessment: A Framework for Evaluating Transit Riders' Opinions about Quality of Service," *Public Transport* 10: 2 (2018) 363–377.

W. Kim, K. Hyun, G. Zhang, and A. Giarrusso, "Social Media Analysis for Transit Assessment" (2019) <https://ctedd.uta.edu/wp-content/uploads/2020/01/Kim_final.pdf> Accessed March 30, 2020.

J. Lee, H., Gao, and K. Goulias, "Can Twitter Data Be Used to Validate Travel Demand Models?" Proceedings of the 95th Annual Transportation Research Board Meeting (Washington, DC, January 2016).

Y. Li, and Y. Liu, "Application of Opinion Mining to Improve Public Satisfaction Based on Comments Regarding Public Transportation Services on Review Sites" paper presented at Transportation Research Board 98th Annual Meeting (Washington, DC, 2019).

M. Manville and A. Levine, "What Motivates Public Support for Public Transit?" *Transportation Research Part A: Policy and Practice* 118 (2018) 567–580.

H. Masood and A. Idris, "An Expert Opinion Survey for Transit Stop Planning," Proceedings of the 97th Annual Transportation Research Board Meeting (Washington, DC, January 2018).

M. Nik-Bakht, and T. El-diraby. "Communities of Interest—Interest of Communities: Social and Semantic Analysis of Communities in Infrastructure Discussion Networks," *Computer-Aided Civil and Infrastructure Engineering* 31: 1 (2016): 34–49.

B. Pender, G. Currie, A. Delbosc, and N. Shiwakoti, "Social Media Use during Unplanned Transit Network Disruptions: A Review of Literature," *Transport Reviews* 34: 4 (2014) 501–521.

P. Pimpale, A. Panangadan, and L. Abellera, "Analyzing Spread of Influence in Social Networks for Transportation Applications," Proceedings of the IEEE 8th Annual Computing and Communication Workshop and Conference (Las Vegas, Nevada, January 2018).

R. Plutchik, *The Emotions*, revised ed. (Lanham, MD: University Press of America, 1991).

B. Qi and A. Costin, "Investigation of the Influence of Twitter User Habits on Sentiment of Their Opinions towards Transportation Services," ASCE International Conference on Computing in Civil Engineering (Atlanta, GA, 2019).

T. Rinker, "sentimentr: Calculate Text Polarity Sentiment at the Sentence Level" (2019) https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf> Accessed March 30, 2020.

R. Sarker, I., Kaplan, M. Anderson, S. Haustein, M. Mailer, and H. Timmermans, "Obtaining Transit Information from Users of a Collaborative Transit App: Platform-Based and Individual-Related Motivators," *Transportation Research Part C: Emerging Technologies* 102 (2019) 173–188.

L. Schweitzer, "Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter," *Journal of the American Planning Association* 80: 3 (2014) 218–238.

W. Shen, W. Xiao, and X. Wang, "Passenger Satisfaction Evaluation Model for Urban Rail Transit: A Structural Equation Modeling Based on Partial Least Squares," *Transport Policy* 46 (2016) 20–31.

TransLoc Marketing, "Top 8 Reasons People Give Up on Public Transit" (June 18, 2013) <https://blog.transloc.com/blog/top-8-reasons-people-give-up-on-public-transit> Accessed March 30, 2020.

J. Wijffels, "udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit," R package version 0.8.2 (2019) <https://cran.r-project.org/web/packages/udpipe/index.html> Accessed March 30, 2020.

B. Wu and A. O. Idris, "Measuring and Visualizing Transit Customers' Satisfaction Using Twitter Data," Proceedings of the 97th Annual Transportation Research Board Meeting (Washington, DC, January 2018).

M. Yeomans, A. Kantor, and D. Tingley, "Politeness: Detecting Politeness Features in Text," R package version 0.3.2 (2019) <https://cran.r-project.org/web/packages/politeness/index.html> Accessed March 30, 2020.