Vision Zero Hashtags in Social Media: Understanding End-User Needs from Natural Language Processing

Subasish Das, Ph.D.
(Corresponding Author)
Associate Transportation Researcher, Texas A&M Transportation Institute
1111 RELLIS Parkway, Room 4414, Bryan, TX 77807
Email: s-das@tti.tamu.edu

Kartikeya Jha
Research Associate, Texas A&M Transportation Institute
1111 RELLIS Parkway, Room 2238, Bryan, TX 77807
Tel: 979-317-2472, Email: jha_k@tamu.edu

Anandi Dutta, Ph.D.
Senior Lecturer, Computer Science and Engineering
Ohio State University, 2015 Neil Ave, Columbus, OH 43210
Email: dutta.34@osu.edu

TOTAL WORDS: 6,174 words

5,424 words = text (including abstract and references)

750=3 table

Submitted on August 1, 2019

ABSTRACT

Vision Zero (VZ) is a global traffic safety policy which promotes road safety by advocating the shared responsibility of designers, policymakers, and road users to work toward the goal of nullifying preventable fatalities and severe injuries caused by man-made errors. This paper investigates the use of the hashtag #VisionZero on Twitter, a popular microblogging platform, to understand public needs and requirements, identify innovative and context-sensitive solutions, and facilitate the exchange of ideas and best practices. To examine the relevance of needs for several different focus areas within VZ, textual content from about 32,000 unique tweets was analyzed using opinion mining, network cluster analysis, and exploratory visual analyses to identify the significance, frequency, and patterns of word strings. Results reveal interesting trends, contexts, and patterns with regard to user reactions to road safety needs, issues, and solutions. This research can help decision-makers in identifying and anticipating the most relevant demands and pressing requirements for different individual areas of interest. Several U.S. cities, Metropolitan Planning Organizations (MPOs), and states have embraced the Vision Zero strategy. This paves the way for a bigger role for the VZ approach in global road safety efforts aimed at achieving results set forth in the Decade of Action for Road Safety in the coming years.

Keywords: traffic safety, Twitter, data mining, interpretable machine learning.

INTRODUCTION

Vision Zero (VZ), regarded globally as an innovative road safety policy, is a strategy with the purpose to eliminate all traffic fatalities and severe injuries by improving safe, healthy, and equitable mobility for all. The fundamental idea of driving this approach is that even though not all road crashes can be entirely prevented, almost all severe injuries and fatalities can be avoided by following safe road design principles (*1*,*2*). First implemented in Sweden in the 1990s, VZ has been gaining success in U.S. cities after being proven successful in Europe. Approximately 40,000 people are killed in U.S. roadways every year. The associated loss for a traffic fatality includes personal economic costs, emotional trauma to those suffering, long-term healthcare costs, and significant taxpayer spending on emergency response and long-term healthcare costs (*3*). There is a need for sharing the information that will not only raises awareness about VZ but also causes institutions and individuals to change their behavior on the streets.

In order to successfully make an effort for VZ, it is important to note that strategic communications are essential. To ensure its effectiveness, the development of a targeted, data-driven approach in spreading the messages of VZ is crucial. A popular microblogging platform, called Twitter, generates approximately 500 million tweets (text-based messages with a 280-character limit) per day. Many individuals and agencies share messages about VZ on Twitter. Understanding the patterns and trends of the messages and texts can aid in the comprehension of the patterns of public engagement. This paper investigates the use of the hashtag #VisionZero on Twitter to explore: 1) topics and their relevance to VZ goals and concepts, 2) opinion mining of the associated texts, 3) the trends and contexts of the associated textual content, and 4) need mining to interpret public needs and requirements.

The structure of this paper is as follows. Initially, the contexts of VZ and associated literature review are provided. Then, the data collection and exploratory text mining are introduced. In the next sections, need mining and emotion measures are described. Finally, the results of this evaluation are shown followed by the final conclusions and discussions.

RESEARCH CONTEXT AND EARLIER WORK

VZ is not just a tagline, slogan, or even just a program but a fundamentally unique approach toward traffic safety. With the ethical belief that people possess the right to move safely within their own communities, the system designers of VZ and policymakers share the obligation to provide safe travel systems. The VZ framework argues that design deficiencies in the present road design philosophy are the main cause of the global road safety crisis, underlining the man-made nature of the issue in the design framework. To reduce the severity of crashes, the policymakers and system designers are expected to improve the roadway environment, policies (such as speed management), as well as other related systems. This policy is different from other safety approaches with regard to problem formulation, requirements for the safety of different road users, views on responsibility for safety from various stakeholders, and attitudes toward the evolving demands of road users to traffic safety *(4)*. VZ is a significant deviation from the existing approaches in two major ways:

- The road system and related policies should be constructed to ensure severe injuries or fatalities do not occur because of human errors.
- VZ recognizes that many variables contribute to safe mobility -- including technology, roadway design, user behavior, vehicle operating speeds, and policies -- and sets clear objectives to achieve the shared goal of eliminating fatalities and severe injuries (*5*).

The three major priorities of VZ are 1) managing speed, 2) centering equity, and 3) engaging communities. Cities which have adopted the VZ approach have experienced success in reducing the hardship of traffic injuries. For example, "the latest VZ report from New York indicates a 28% reduction in the city's traffic fatalities and a 45% decline in pedestrian fatalities since the launch of their program in 2014" (*6*). New York City established an annual VZ Fleet Safety Forum for municipal fleet operators and managers. The Forum answers questions and educates the on-the-ground operators of the large vehicles and the public about the safety improvements (*7*).

Multiple VZ cities have enforced effective plans with regard to speed reduction. Recently, Seattle succeeded in lowering neighborhood street speeds to 20 mph and arterial road speeds to 25 mph. Similarly, Boston was also successful in decreasing local speed limits from 30 to 25 mph (*8*). Additionally, Washington D.C. has plans to add 100 new cameras around schools where the speed limit has been reduced to 15 mph (*9*), and Portland's Department of Transportation recently won the approval of the state to permanently reduce the speed limit from 35 to 30 mph on one of its High Crash Network streets (*10*).

The U.S. currently has no federal regulations that render the use of side guards. The policy advancement regarding side guards has been initiated locally. Cities that have created policies requiring side guards on trucks in certain circumstances include Boston, Portland, Washington D.C., New York City, and Cambridge. (*7*).

San Francisco's VZ Action Strategy (*11*) advocated a transformative policy plan that prioritizes the strategies that make the largest impact on safety. The agenda laid out the similarities of VZ goals aligned with other top city goals and policies "that prioritize walking, biking and improved transit while reducing driving and vehicle miles." Emerging VZ data-driven approaches focus on actions that are "proven to reduce instances of vehicle conflicts." Lastly, they intend to make the work of VZ accountable and measurable. For example, the percentage of safety treatments is measured and improved in Communities of Concern.

Seattle aims to eradicate traffic fatalities by 2030 and has highlighted three areas of focus: education and public engagement; street design, policies, and regulation; and enforcement. According to the mayor of Seattle, Ed Murray, one of the key components in decreasing motorist speed is to comprehend the life-saving implications of driving within an acceptable range (*12*).

The Bicycle Transportation Alliance and Oregon Walk organized a report in which they emphasized the issues and statistics on traffic crashes and injuries data in the Portland Metropolitan Region. The report proposed solutions from developments in state policy to shifts in street design practice. One of the most significant solutions includes engaging the region's diverse communities with large non-English speaking populations to ensure that community education and engagement strategies are culturally suitable, and they are able to influence any planned adjustments to their neighborhood streets and to any new policies and laws. Other suggested solutions include enforcing road diets on dangerous streets and high-risk corridors which will reduce travel lanes, including pedestrian infrastructure and safe bicycling, and eliminating dangerous and illegal speeding (below the 85% observed speed). In Massachusetts, the Boston City Council authorized a citywide truck side guard ordinance in 2014 that mandates "all city-owned trucks purchased after July 1, 2014, as well as city-contracted vendors to be equipped with the following: side guards, convex mirrors, cross-over mirrors, and blind-spot awareness decals. All contracted vehicles must be approved before any work or services are provided" (*7*).

The number of cities adopting the VZ approach has more than doubled, and the interest in adopting evidence-based, action-driven policies to improve road safety has increased drastically from large and small communities to big, coastal cities across the country (*13*). VZ is not limited to big cities only, as smaller cities have also implemented innovative action plans. For example, "Fremont, California, has been an early VZ adaptor and implemented some highly effective VZ projects. First, Fremont converted its entire streetlight system (16,000 lights) to brighter, white LEDs for improved nighttime visibility. This action dramatically reduced the number of major crashes occurring at nighttime from 47% to 29%. Second, they found that pedestrians are more likely to being hit at signalized intersections because they lacked information regarding available pedestrian crossing time. Therefore, they increased the number of countdown signals from 23% to 100% of the City's 220 traffic signals. Third, as a low-cost way to redesign streets for lower speeds and greater safety, Fremont is using its pavement maintenance program to restripe streets with narrower lanes (10 feet), including buffered bike lanes, and paint high visibility or continental style crosswalks" (*14*). To reduce speed limits and allow safety cameras in San Jose, reports state that large investments should be devoted to enforcing policy changes as well as toward promoting automakers who have already created self-driving technology and crash avoidance that can be included in the solution (*15*).

Vision Zero Studies

In 2009, Johansson analyzed the philosophy of safety in the current road and streets designs. After providing background for the origin of the philosophy, the proposition for a new approach to road and street design was presented (*3*). These design principles contributed to the framework design for safer systems which have built-in thresholds for the tolerable losses of human health for all predicted crashes *(16)*. Similarly, the application of strategies influenced by VZ philosophies and principles to the modern road and car designs has been evaluated by Kim et al. *(17)*. Although rising interest in VZ and its application has been expressed by multiple U.S. states, very little literature exists to detail the actions and purpose of VZ to eradicate fatalities over the next decade *(18)*.

As a result of the growing interest in VZ as of 2001, approximately 30 U.S. states have adopted programs geared toward zero traffic fatalities under different names such as Target Zero, VZ, or Toward Zero Deaths (TZD) in cities and states like Seattle, Washington; Portland, Oregon; San Francisco, San Jose, San Mateo, and San Diego, California; Washington, D.C.; New York City; and Boston, Massachusetts. In 2003, Minnesota adopted the zero fatality goal after the adoption support of Washington State in 2000.

To boost public knowledge regarding VZ and its potential applications, Fleisher et al. *(18)* presented a framework for cities' use to benchmark their efforts relative to other jurisdictions, identify efficient strategies, and extend to other cities and countries' officials who pursue additional information. Solutions were also provided to relate their findings with the implementations of VZ in U.S. cities. The strategies, efforts, development, implementation time frames, areas of focus, outreach, and communications of the National Strategy on Highway Safety were presented by Sores and Costales (*19*) with the objective of implementation of practice in the United States. Rosencrantz et al. (*20*) discussed the experiences with VZ and contextualized the criticism VZ faced.

While some U.S. states target a year to achieve zero fatalities (e.g., Washington State), some states do not establish a specific end date. "Minnesota, while lacking a final success date, does have interim goals. Sweden's program, like Minnesota, has no official end date and therefore it will never be too late to achieve the goal." Although several of the state TZD programs are emerging and relatively new, four state programs have sufficient crash data to analyze the impact of the programs in Minnesota, Idaho, Utah, and Washington. Munnich et al. (*21*) analyzed the effectiveness of VZ Programs. The implementation of TZD programs has reduced road fatality rates. Because these rates vary from state to state, any new program would require a necessary time investment to reach its full potential. Each state has various degrees of the temporal effect of its own TZD program, so the average impact is going to become more evident over time.

Cushing et al. (*22*) reviewed the current VZ policies and linked literature to identify key factors of VZ policy in Sweden and provide evidence-based suggestions for the implementation of similar policies in the United States. The study suggested that infrastructure design should remain a priority in the United States VZ plans. However, incorporating supplemental initiatives can also improve road safety culture.

DATA DESCRIPTION

Data Collection

With approximately 500 million tweets per day, Twitter provides real-time big textual contents 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 in real-time. Some of the information and unfiltered opinions can be very sensitive in various aspects. Twitter generates a huge amount of textual content daily. This research used the open-source R software package 'twitteR' to collect the data *(23)*. The data collection was done using the Open Authorization (OAuth), an authentication mechanism that allows applications to provide client functionality to a web service without granting an end user's credentials to the client itself. Authentication is required for all Twitter-related data collection. Table 1 provides definitions for relevant Twitter terms.

Table 1 Definitions of Twitter Terms

Analytics	Definition
Tweet	User or handle post (limited to 280 characters)
Engagements	Interactions in Twitter
Handle	Username in Twitter (e.g., jen_keesmaat)
Impressions	Times Twitter user were shown a tweet in the timeline or search results
Likes	Number of Twitter user who "liked" the tweet
Retweets	Number of times Twitter user reshared the tweet
Replies	Number of times Twitter user replied to the tweet
Embedded media clicks	Clicks to view a photo or video in the tweet
Followers	Times someone began following the account directly from the tweet
Hashtag clicks	Clicks on hashtag(s) (#) in the tweet

To collect relevant data, several keywords have been used during the data collection process. The keywords include *'visionzero' and 'zerovision.'* The dataset was collected for four months in 2019 (March 12, 2019, to July 16, 2019). A total of 32,000 unique tweets were collected. These tweets are retweeted 69,450 times. The tweets are associated with approximately 18,283 Twitter handles. Figure 1 shows the top four tweets with the maximum number of retweets during the data collection period. The tweet with the largest number of retweets with hashtag #VisionZero has over 2,000 retweets which were tweeted by @ETAservicesltd.

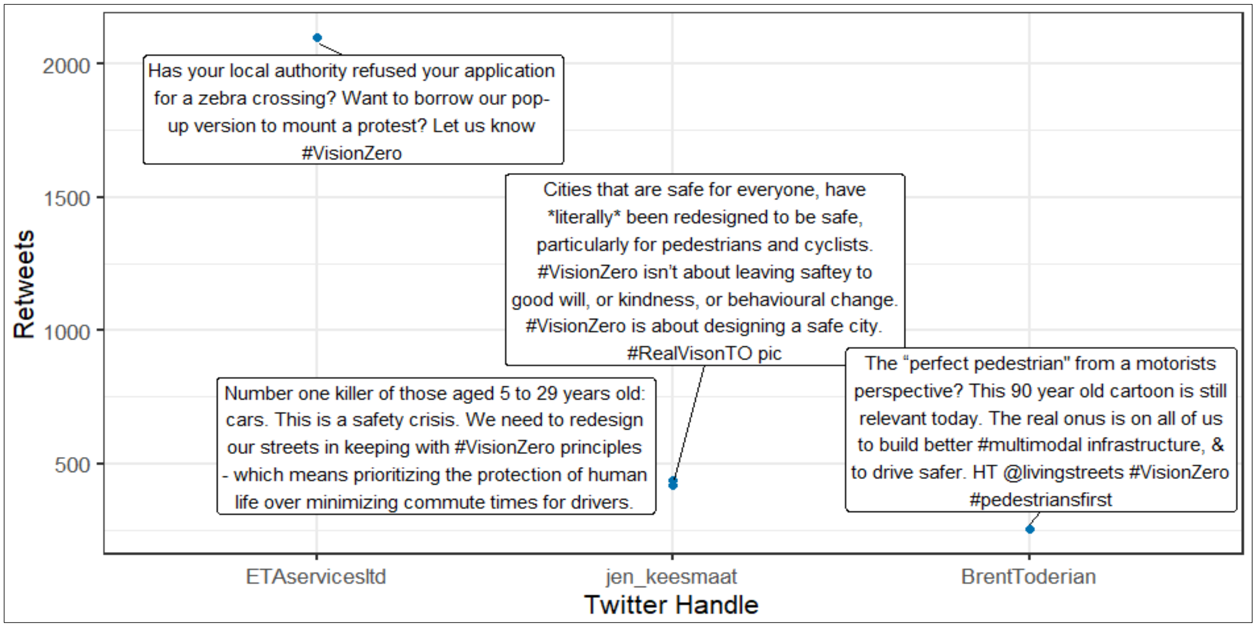


Figure 1 Top four #VisionZero tweets with the highest number of retweets.

METHODOLOGY

Exploratory Text Mining

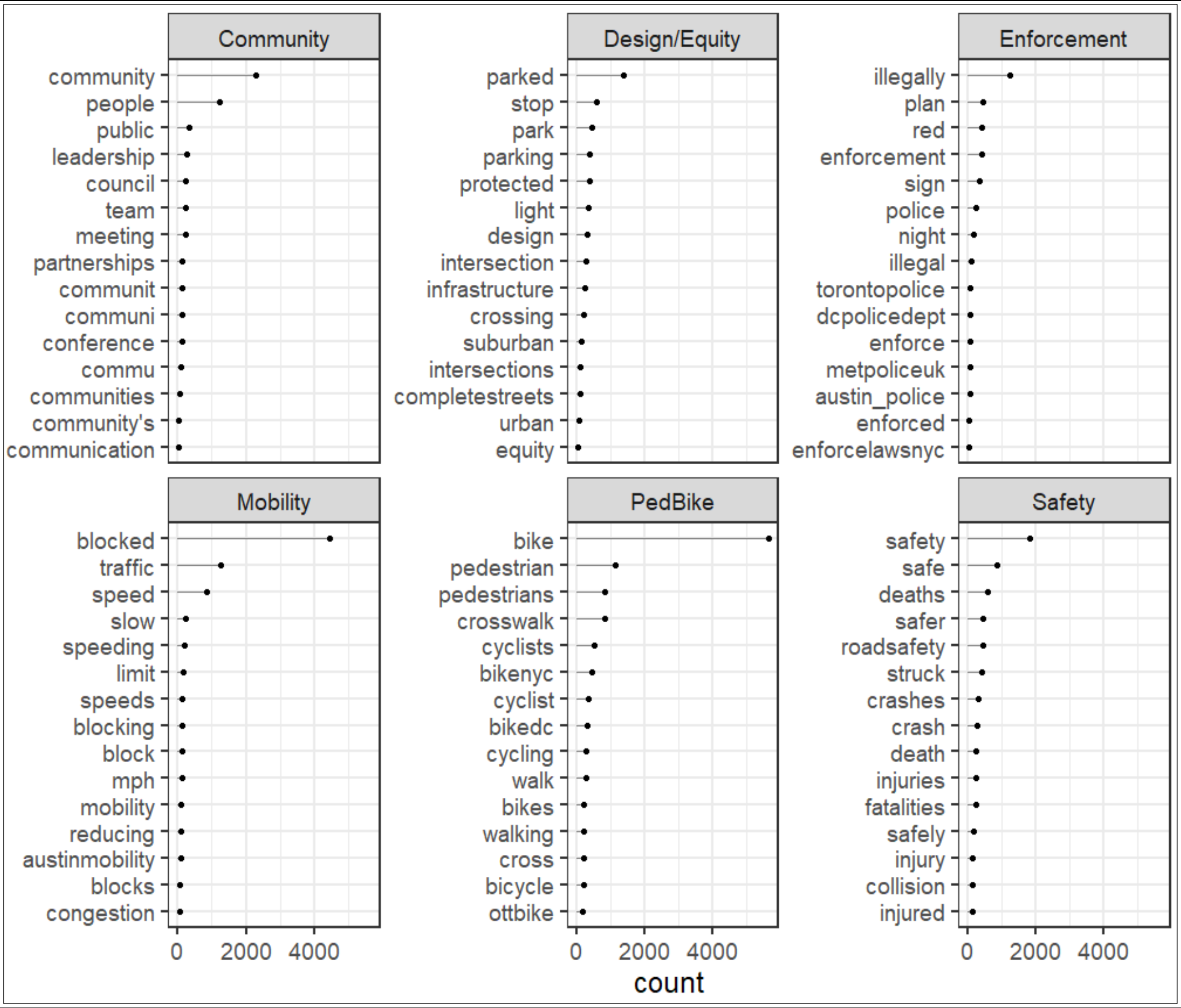
Keyword Extractions for VZ Key Focus Areas

The keywords associated with VZ tweets were identified for different key focus areas within the VZ approach. Figure 2a shows these keywords (also known as unigram) along with their frequency of occurrence in tweets during the study period. The observations map well with the fundamental concepts of VZ. For example, one of the basic premises behind VZ is the shared responsibility of road safety between designers, policymakers and road users *(2-5,24)*. These VZ concepts and experiences are reflected in the frequency of keywords for each focus area. Keywords such as community, speed, parking, bike, pedestrians, design, and safety rank high on the list of keywords tweeted.

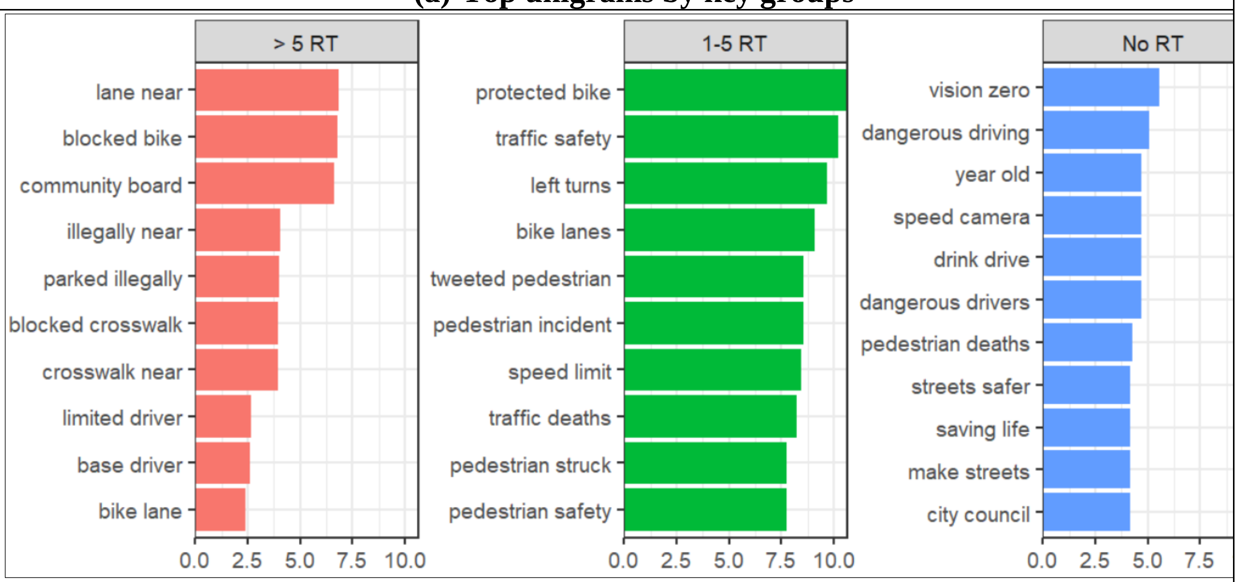
Odds Ratio

The odds ratio quantifies the strength of association between two events and is considered a good statistical tool for finding associations between categorical variables. It represents the ratio of the odds of an event A in the presence of another event B and the odds of event A in the absence of event B. A higher number indicates stronger association. The *odds* of word *w* in group i's usage can be defined as:

$$Odds_w^{(i)} = p_w^{(i)}/(1 - p_w^{(i)}) \quad (1)$$



(a) Top unigrams by key groups



Log odds ratio

Figure 2 Unigram and bigram analysis plots.

where, $p_w^{(i)} = \frac{t_w^{(i)}}{n_k^{(i)}}$. The term $t_w^{(i)}$ denotes the vector of word frequencies from documents of class *i* for topic *k*. The odds ratio between two groups can be written as $\theta_w^{(g1-g2)} = Odds_w^{(g1)} / Odds_w^{(g2)}$. This is generally presented for single words in isolation or as a metric for ranking keywords. Figure 2b shows the log odds ratios of the top bigrams (a sequence of two consecutive words), the continuous sequence of two words from a document, from documents developed based on the number of retweets (> 5 RT: five or more retweets, 1-5 RT: one to five retweets, No RT: no retweets). For example, the bigram 'blocked bike' is 6.45 times more likely to be present in a tweet with a high number of retweets (> 5RT) than in a tweet with fewer retweets.

Network Visualizations

In Figure 3, the network plot displays the network patterns of the Twitter handles that are associated with VZ related tweets. Using Gephi 0.9.2, the research team created network plots to analyze the network of users within the gathered tweets *(26)*. First, the research team used the R software to produce a GDF file, with link-in and link-out counts (the number of times each account is retweeted and retweeting others) as a characteristic. Then, the GDF file was imported to Gephi. Using the ForceAtlas algorithm, the research team arranged a network visualization to group the nodes with the similar connection. The node sizes are proportional to color and link-in counts by different nodes.

The graph represents the relationship between retweets associated with #VisionZero. According to the clusters shown in the graph, the nodes can be separated into five major clusters, which also correspond to the geolocations. The clusters with @NYC_DOT and @NYPD19Pct represent the U.S. VZ interests. Canada has two large clusters representing Ottawa and Toronto. The graph indicates that Canada pays the most attention to the VZ Program. There are some users who have a large influence on the VZ program in Canada. The handles include @BrentToderian, @jen_keesmaat, @cmckenney (Ottawa City Councillor), and @RZaichkowski (blogger for Two-Wheeled Politics bike). The graph on the left shows the retweeting cluster represented by @MPSRTPC (media for MPS Roads & Transport Policing Command), @SuptAndyCox (specialist for VZ Program), and @MetCycleCops (officers for cycle safety in London). The right portion of the graph comes mostly from Germany. The handle @ADFC_Berlin is the user with the most retweeted counts in Germany and focuses mostly on bike safety.

The observations fall in line with a growing consensus in North America and around the world for a holistic, safe, and user-friendly transportation system *(24)*. There is increasing agreement for allowance of safe interactions between motorized and non-motorized modes of transportation infrastructure to encourage the use of walking and biking for low mileage trips, carpooling, and shared mobility for peak period travel, etc.

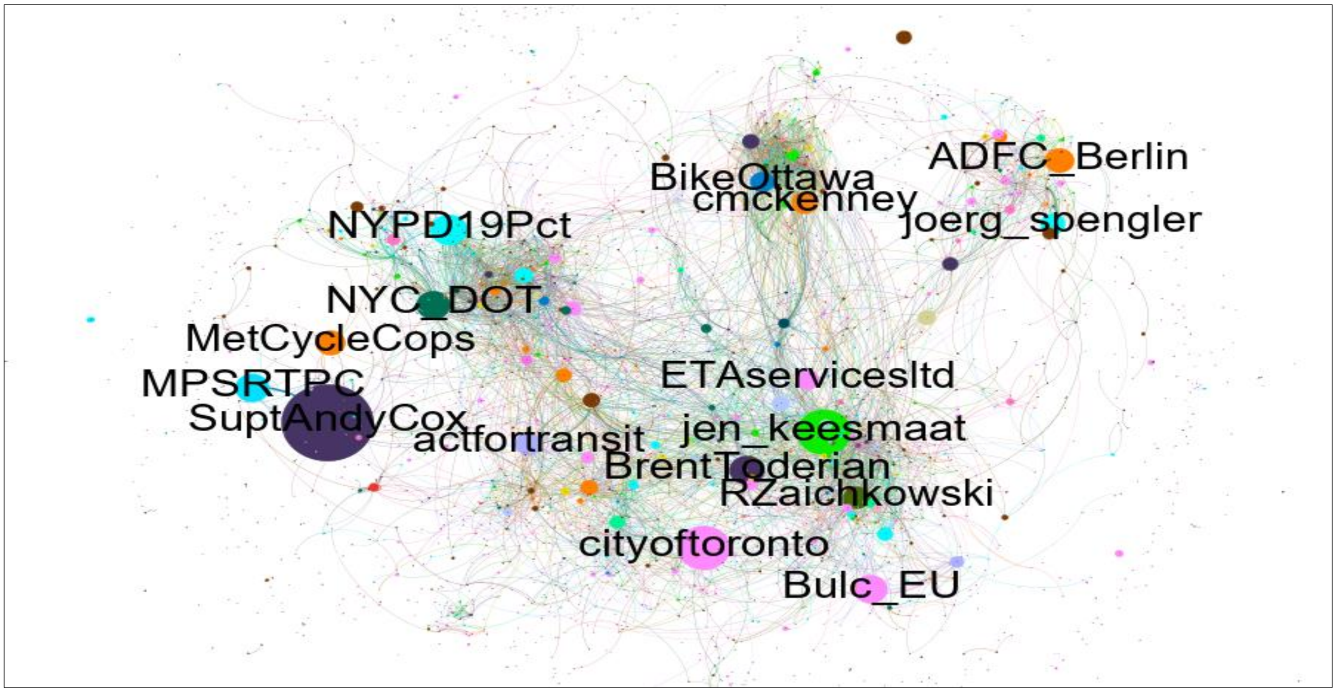


Figure 3 Network pattern of Twitter handles for VZ tweets.

Vision Zero Need Mining

Identifying the needs and demands of VZ hashtag users is an innovative way to design new market-driven services. This study evaluates the use of Twitter data as a source for identifying customer needs. One of its advantages is the real-time aggregation of needs. These needs are expressed autonomously by the customer, for example, in a moment of dissatisfaction or inconvenience with the underlying demand for a certain agenda or solution. The need related

Table 2 Interactive Table Showing #VisionZero Tweets on Needs (Source: https://rpubs.com/subasish/514501)

	screenName	text	word	retweetCount
1144	DR_6B09	RT @schlthss: #VisionZero demands infrastructure changes to ensure people are not killed on dangerous by design roads as Dave was yesterday□	demands	35
2	linzcollins	RT @bostonbikeunion: In 2017 #GoBoston2030 identified Mass Ave as a #VisionZero priority corridor for safety improvements. Changes were mad□	improvements	24
3874	pwbnyc	RT @kpeinhardt: Tonight's powerful reminder from the #NYCBikeFamily: People over parking. #FixOurStreets #VZEmergency #VisionZero https://t□	fixourstreets	21
3873	kpeinhardt	Tonight's powerful reminder from the #NYCBikeFamily: People over parking. #FixOurStreets #VZEmergency #VisionZero https://t.co/OPxmaw8qQ3	fixourstreets	21
434	grahamprojects	RT @NYC_DOT: Since the launch of #VisionZero in 2014, DOT has pioneered and launched many new road safety treatments. Let□s take a look a□	treatments	17

words used in this filtering include *need, want, wish, feature, ask, improve, idea, upgrade, support, problem, issue, solution, help, fix, complain, fail, solve, countermeasure, treatment, require, demand, urgent, urgency,* and *please*. The final dataset contains 1,041 unique keywords with a total frequency of 52,409. The need related terms appeared 3,738 times (7.14% of final keywords) in the collected tweets. The number of need related tweets is 4,089 (proximately 13% of the collected tweets). The authors developed an interactive web link for the tweets prepared for further needs assessment *(27)* as shown in Table 2.

Term Frequency Inverse Document Frequency (TF-IDF)

In information retrieval, TF-IDF has been widely used to differentiate between documents (one single tweet can be considered as a document and group of tweets can be considered as corpus) by estimating how relevant their contents are to a set of terms in a search string. It combines two different weighting parameters to determine the relevance of the document: TF and IDF. The concept of TF-IDF has been taken from Silge and Robinson *(28).* In TF-IDF, terms are viewed as having different levels of importance- some terms are weighted more while others are weighted less. TF can be denoted as $tf(t, d)$, which indicates the number of occurrences of the term (t) in the document (d). The IDF of a term (t) can be narrated as:

$$idf(t, D) = log \frac{|D|}{df(t, D)} \quad (2)$$

where $|D|$ is total frequencies of documents in the corpus D, and $df(t, D)$ is the measure that indicates the frequency of documents that contain word or term t:

$$df(t, D) = |\{d \in D: t \in d\}| \quad (3)$$

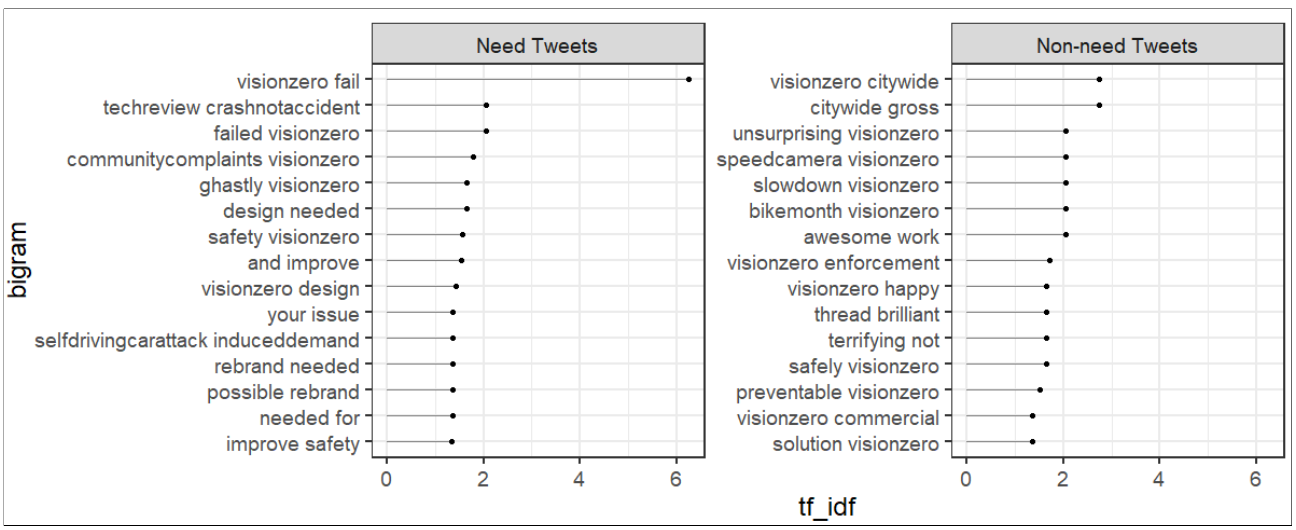


Figure 4 Top bigrams with high tf-idf for need and non-need tweets.

The parameter IDF reduces the weightage of terms that often appear in all documents and provides additional weight to terms that are only found in a few documents to differentiate between documents. The TF-IDF score for a term, t, is the product of the IDF and TF scores. For a string consisting of the set of terms q, the TF-IDF score of document d in corpus D is *(28):*

$$TF - IDF(q, D) = \sum_{t \in q} tf(t, d) \times idf(t, D) \quad (4)$$

Figure 4 illustrates the top 15 keywords by the tweet types. The keywords are sorted based on the TF-IDF values. One common assertion is that the need tweets are associated mostly with problems, failures, and other need relevant word groups.

Difference between Sentiment Scores

The research team used the concept of sentiment scores from the open-source R software package, '*sentimentr*.' *(29).* To determine the sentiment scores for this study, the augmented dictionary method was used. This study considers valence shifters when determining these values, which is more efficient than a simple lookup dictionary approach. This study provides a brief overview of the concept of sentiment score, which is mostly based on Rinker *(29).*

An equation is used to assign the polarity of each sentence a value in the algorithm. Using a sentiment dictionary, it tags polarized words first. Each paragraph, composed of sentences ($\mathrm{p_i} = \{\mathrm{s_1}, \mathrm{s_2}, \ldots, \mathrm{s_n}\}$), is fragmented into several element sentences ($s_i, j = \{\mathrm{w_1}, \mathrm{w_2}, \ldots, \mathrm{w_n}\}$) where w is the words within sentences. Then, each sentence (s_j) is divided into an ordered bag of words. The words, as an i, j, k, are notated as $w_{i,j,k}$. cw denotes the pause words, or comma words.

To assign a sentiment value, the words in each sentence ($w_{i,j,k}$) are inspected and compared to a dictionary of polarized words. Negative words are labeled with a -1, positive words are labeled with a $+1$. Based on the used senti-lexicon, the sentiments are tagged as either positive and negative. The word with polarization measures (pw) will form a polar cluster ($c_{i,j,l}$), a subset of the sentence ($c_{i,j,l} \subseteq s_i, j$). The polarized context cluster ($c_{i,j,l}$) of words is pulled from around the pw and the n1 words before and n2 words after are considered as valence shifters. The cluster can be represented as ($c_{i,j,l} = \{pw_{i,j,k-nb,\ldots,}\, pw_{i,j,k}\, , pw_{i,j,k-na}\}$), where nb & na are the parameters before and after set by the user. In this polarized context cluster, the words are labeled as negator (${w_{i,j,k}}^n$), neutral (${w_{i,j,k}}^0$), amplifier (${w_{i,j,k}}^a$), or de-amplifier (${w_{i,j,k}}^d$). Although neutral words contribute to the word count (n), they hold no value in the equation.

Figure 5 shows the emotional propensity of the VZ related tweets for different emotion groups. From the emotion measures, a preponderance of negative emotion is dominant. Given the hashtag is associated with traffic crashes and severities, it is obvious that the VZ tweets are associated with negative connotations. It is often debated that tweets associated with negative emotion could provide more feedback than positive tweets. Additionally, it is important to note that the emotions and sentiments are measured based on the conventional sentiment lexicons. There is a need for transportation science-related sentiment lexicon that can measure more domain-specific emotions and sentiments. While the current study suggests this development as an area for future improvement, as a whole, it can be said that emotion mining measures can provide evidence of the overall contexts associated with VZ tweets.

Table 3 Summary Statistics and Sentiment Scores for Emotions in VZ Tweets

Emotion Type	Total Emotions (Need Tweets)	Total Emotions (Non-need Tweets)	Average Emotions (Need Tweets)	Average Emotions (Non-need Tweets)
anger	538	321	0.009194	0.005811
anger_negated	15	4	0.000237	0.000066
anticipation	862	651	0.014936	0.012067
anticipation_negated	17	10	0.000265	0.000183
disgust	272	168	0.004824	0.002930
disgust_negated	6	2	0.000098	0.000035
fear	811	556	0.014166	0.009979
fear_negated	19	15	0.000309	0.000259
joy	749	412	0.013016	0.007506
joy_negated	7	9	0.000102	0.000152
sadness	649	435	0.011428	0.007860
sadness_negated	19	15	0.000306	0.000241
surprise	296	211	0.004960	0.003926
surprise_negated	10	13	0.000155	0.000208
trust	1238	756	0.021412	0.013621
trust_negated	12	16	0.000202	0.000279

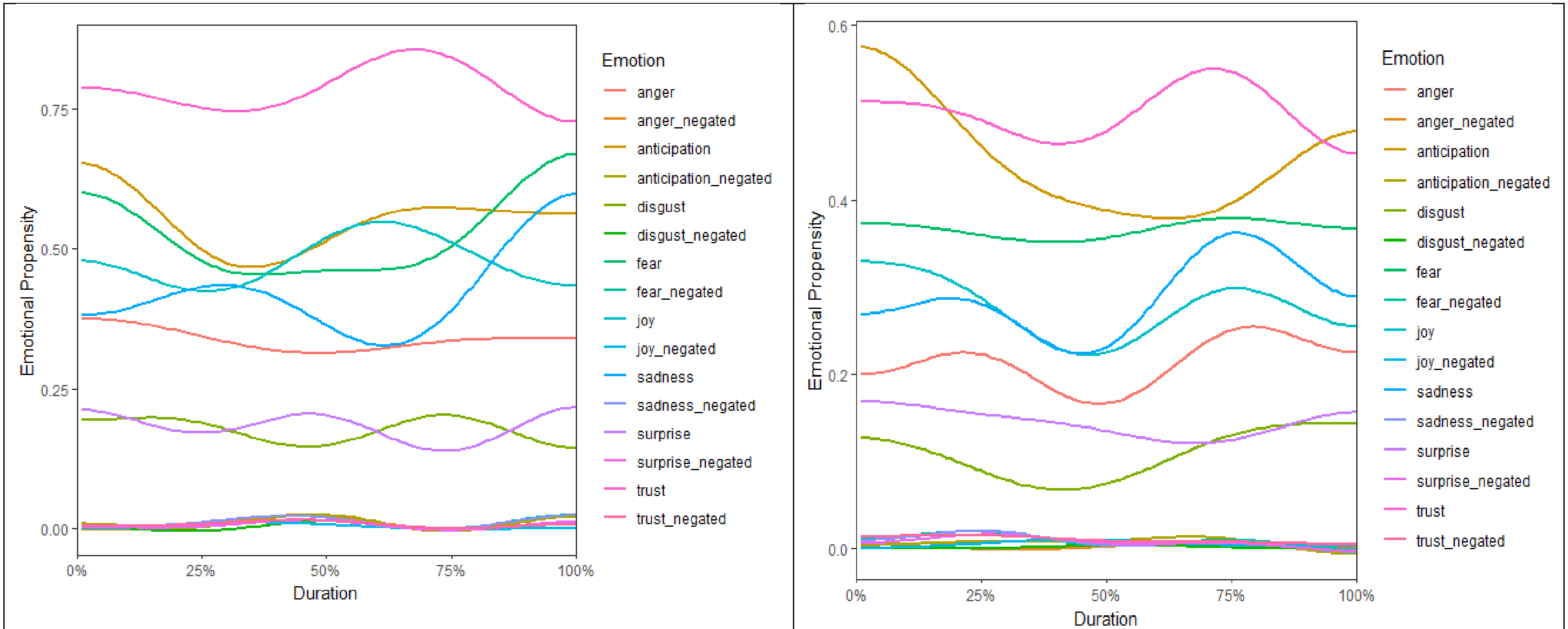
Valence Shift Word Measures

Dodds and Danforth *(30)* emphasized the importance of 'Valence Shift Word Graph' in their study. To categorize text as negative or positive, sentiment analysis is often applied. Indeed, the value added from sentiment analysis would be restricted if it were the only use case. However, sentiment analysis is utilized as a lens that permits the user to grasp how the overall context is shaped by emotive words in a text. Before examining the difference in sentiment scores between two texts, each word must be analyzed to find its individual contribution. The final and most important step is to examine the words themselves that are ordered by their individual contribution.

First, the contribution of each word must be found individually to make this possible. Starting with the ANEW sentiment dictionary *(31)* and two texts which are labeled reference and comparison, the user takes the difference of their sentiment scores $s_{ANEW}^{(comp)}$ and $s_{ANEW}^{(ref)}$, rearrange terms, and arrive at the following *(30):*

$$s_{ANEW}^{(comp)} - s_{ANEW}^{(ref)} = \sum_{w \in ANEW} \underbrace{\left[h_{ANEW}(w) - h_{ANEW}^{(ref)}\right]}_{+/-} \underbrace{\left[p^{comp}(w) - p^{ref}(w)\right]}_{\uparrow/\downarrow}$$

According to its change in usage frequency (↑ / ↓) and its sentiment relative to the reference text (+/–), each word w in the summation contributes to the sentiment difference between the texts. To present the information in a manner that is accessible to the users, word shift graphs are used. Figure 6 be interpreted by the following interpretations:

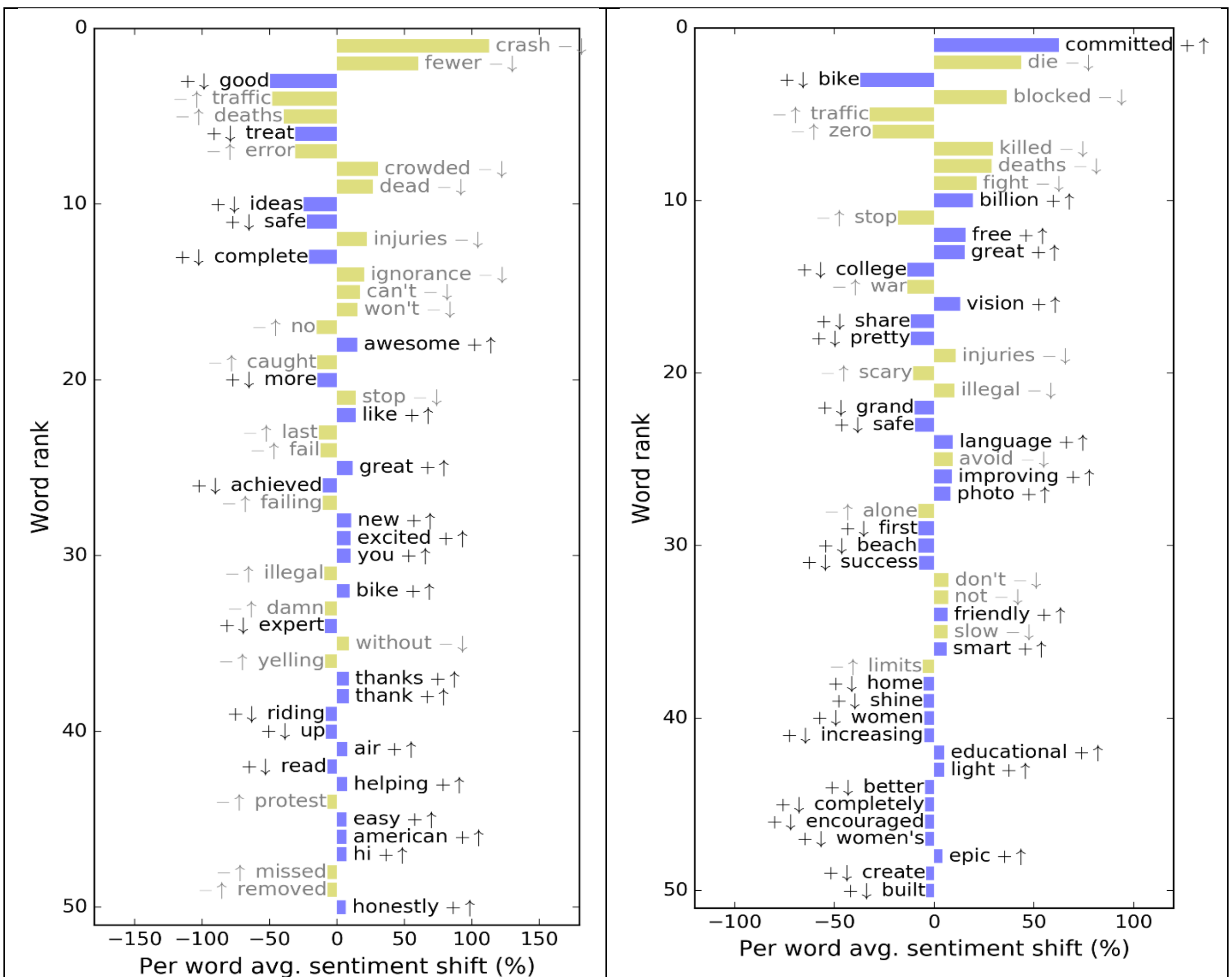


(a) Need related Tweets | **(b) Non-need related Tweets**

Figure 5 Emotion propensity plots.

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

The x-axis of the valence shift plots shows per word average sentiment shift by word (and the rank in y-axis is dependent on this value; for example, word with highest shift will be ranked as 1). Crash and fewer are the top two words in the need tweets. Both words are on the right-hand side (in yellow bar), which indicates that these two negative words are less used. On the other hand, the word 'committed' (considered as positive based on the generalized sentiment lexicon) has the highest shift (located on the right-hand side with purple bar) which indicates that this term was more used in the 'non- need' related tweets. All the words in this plot can be interpreted similarly by the readers based on their interests.



(a) Need related tweets **(b) Non-need related tweets**

Figure 6 Valence shift word plots.

CONCLUSIONS

VZ, a traffic safety policy which originated in Sweden in 1997, has gained traction across different parts of the world and is being adopted by different state DOTs, MPOs, and city agencies across the U.S. This approach promotes road safety by advocating the shared responsibility of designers, policymakers, and road users to work toward the ultimate goal of nullifying preventable fatalities and severe injuries caused by man-made errors.

Twitter, a popular microblogging platform, generates huge, real-time data that can be tapped to understand the challenges, issues, and needs of road users, identify innovative and context-sensitive solutions, and exchange ideas and best practices. This paper investigates the use of the hashtag #VisionZero on Twitter to explore: 1) the contexts and trends of the associated textual content, 2) topics and their relevance to VZ concepts and goals, 3) opinion mining of the associated texts, and 4) need mining to understand public needs and requirements.

Text content from about 32,000 unique tweets was analyzed using different text mining techniques (need mining and opinion mining, network cluster analysis and exploratory visual analyses) to identify the significance, frequency, and patterns of word strings to examine the relevance of needs for several different focus areas within VZ. Results reveal interesting trends and patterns with regard to user reactions to road safety needs, issues, and solutions. This kind of examination can help policy- and decision-makers in identifying and anticipating the most relevant demands and pressing requirements for different individual areas of interest. The authors also developed an interactive web link for the tweets which can be used for the further needs assessment.

Studies have shown that when large-scale attempts are made to implement VZ approaches and design principles, fatalities have lowered to as much as one-tenth of the initial risk level, particularly in built-up areas and along major rural highways. Previous research concludes that large-scale, alternative, and cost-efficient implementation of these safety principles is feasible and effective. The goal remains to reduce the number of fatalities and serious injuries drastically, achieving the number “zero” figuratively, if at all not literally. Several U.S. cities, Metropolitan Planning Organizations, and states have also embraced the VZ strategy. This paves the way for a bigger role of VZ approach in global road safety efforts aimed at achieving results set forth in the Decade of Action for Road Safety in the coming years *(32)*.

The current study is not without limitation. First, the data collection was only limited to several months. Second, location specific (for example, VZ in the U.S.) analysis was not conducted. For future work, the methodology adopted in this study can be extended to specific locations with large number of archived tweets to potentially further corroborate the findings from this study.

ACKNOWLEDGMENT

The authors appreciate the assistance provided by the students on this project: Bita Maraghehpour, and Ly-Na Tran.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm the contribution to the paper as follows: study conception and design: Subasish Das, Kartikeya Jha, Anandi Dutta; data collection: Subasish Das; analysis and interpretation of results: Subasish Das, Kartikeya Jha, Anandi Dutta; draft manuscript preparation: Subasish Das, Kartikeya Jha, and Anandi Dutta. All authors reviewed the results and approved the final version of the manuscript

REFERENCES

1. Corben, B. F., D. B. Logan, L. Fanciulli, R. Farley, and I. Cameron. Strengthening Road Safety Strategy Development ‘Towards Zero’ 2008–2020 – Western Australia’s Experience Scientific Research on Road Safety Management SWOV Workshop 16 and 17 November 2009. Safety Science, Vol. 48, No. 9, 2010, pp. 1085–1097.
2. Elvebakk, B., and T. Steiro. First Principles, Second Hand: Perceptions and Interpretations of Vision Zero in Norway. Safety Science, Vol. 47, No. 7, 2009, pp. 958–966.
3. Johansson, R. Vision Zero – Implementing a Policy for Traffic Safety. Safety Science, Vol. 47, No. 6, 2009, pp. 826–831.
4. Belin, M. Å., P. Tillgren, and E. Vedung. Vision Zero – a Road Safety Policy Innovation. International Journal of Injury Control and Safety Promotion, Vol. 19, No. 2, 2012, pp. 171–179.
5. Communities Stand Together for Safety, Vision Zero. https://visionzeronetwork.org/communities-stand-together-for-safety-vision-zero/. Accessed Jul. 15, 2019.
6. New York City Department of Transportation. Vision Zero: Year Four Report. New York City, 2018.
7. Case Study: How Can Cities Increase the Safety of Large Vehicles in Urban Areas? https://visionzeronetwork.org/case-study-how-can-cities-increase-the-safety-of-large-vehicles-in-urban-areas/. Accessed: Jul. 15, 2019.
8. Boston Is Set to Lower Its Speed Limit in Jan. - The Boston Globe. https://www.bostonglobe.com/metro/2016/11/30/slow-down-boston-speed-limit-change-takes-effect-jan/FUYH10IQs5SUAc5CoCqpsJ/story.html. Accessed: Jul. 15, 2019.
9. D.C. Plans to Add More Traffic Cameras - The Washington Post. https://www.washingtonpost.com/news/dr-gridlock/wp/2015/12/16/d-c-plans-to-add-100-more-traffic-cameras/?utm_term=.ffefeebe3740&noredirect=on. Accessed: Jul. 15, 2019.
10. PBOT News Release: Portland Gains State OK for Permanent Speed Reduction on Outer Division, a High Crash Corridor. Portland Bureau of Transportation. https://content.govdelivery.com/accounts/ORPortland/bulletins/19d5c81. Accessed: Jul. 15, 2019.
11. San Francisco Raises the Notch for Vision Zero in New Action Strategy: Aiming for Transformative Change. https://visionzeronetwork.org/san-francisco-raises-the-notch-for-vision-zero-in-new-action-strategy-aiming-for-transformative-change/. Accessed: Jul. 15, 2019.
12. From Seattle to Boston: The Momentum for Vision Zero Nationwide. https://visionzeronetwork.org/from-seattle-to-boston-the-momentum-for-vision-zero-nationwide/. Accessed: Jul. 15, 2019.
13. Top 5 Vision Zero Trends of 2016. https://visionzeronetwork.org/top-5-vision-zero-trends-of-2016/. Accessed: Jul. 15, 2019.
14. Vision Zero-Not Just for Big Cities. https://visionzeronetwork.org/vision-zero-not-just-for-big-cities/. Accessed: Jul. 15, 2019.
15. San Jose: Widening the Lens on Vision Zero. https://visionzeronetwork.org/san-jose-widening-the-lens-on-vision-zero/. Accessed: Jul. 15, 2019.

16. Tingvall, C. Vision Zero and The New Paradigm for Road Safety (Emphasizing Critical Lessons for Infrastructure Design). Second UN Stakeholder Forum on Global Road Safety, Geneva, 2007.
17. Kim, E., P. Muennig, and Z. Rosen. Vision Zero: A Toolkit for Road Safety in the Modern Era. Injury Epidemiology, Vol. 4, No. 1, 2017, p. 1.
18. Fleisher, A., M. L. Wier, and M. Hunter. A Vision for Transportation Safety: Framework for Identifying Best Practice Strategies to Advance Vision Zero. Transportation Research Record, Vol. 2582, No. 1, 2016, pp. 72–86.
19. Sorel, T. K., and T. E. Costales. Toward Zero Deaths: A National Strategy for Highway Safety. TR News, No. 282, 2012.
20. Rosencrantz, H., K. Edvardsson, and S. O. Hansson. Vision Zero – Is It Irrational? Transportation Research Part A: Policy and Practice, Vol. 41, No. 6, 2007, pp. 559–567.
21. Munnich, L. W. J., F. Douma, X. Qin, J. D. Thorpe, and K. Wang. Evaluating the Effectiveness of State Toward Zero Deaths Programs. Center for Excellence in Rural Safety, 2012.
22. Cushing, M., J. Hooshmand, B. Pomares, and G. Hotz. Vision Zero in the United States Versus Sweden: Infrastructure Improvement for Cycling Safety. American Journal of Public Health, Vol. 106, No. 12, 2016, pp. 2178–2180.
23. Gentry, J. twitteR: R Based Twitter Client. R package version 1.1.9. 2015. https://CRAN.R-project.org/package=twitteR Accessed: July 2019.
24. Ecola, L., S. W. Popper, R. Silberglitt, and L. Fraade-Blanar. The Road to Zero. Rand Health Quarterly, Vol. 8, No. 2, 2018.
25. Gephi 0.9.2 API Index, https://gephi.org/gephi/0.9.2/apidocs/ Accessed: July 2019.
26. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. Political Analysis. Vol 16(4), 2008, pp. 372-403.
27. Das, S. Needs Related Tweets using #VisionZero. http://rpubs.com/subasish/514501 Accessed: July 2019.
28. Silge, J., and D. Robinson. Text Mining with R: A Tidy Approach. O'Reilly Media, 2017.
29. Rinker, T. sentimentr: Calculate Text Polarity Sentiment, Accessed via https://cran.r-project.org/web/packages/sentimentr/index.html. Accessed: July 2019.
30. Dodds, P, and C. Danforth. Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. Journal of Happiness Studies, 11(4), 2010, pp 441–456.
31. Bradley, M., and P. Lang. Affective norms for English words (anew): Stimuli, instruction manual and affective ratings. Technical report c-1, University of Florida, Gainesville, FL, 1999.
32. The UN Road Safety Collaboration. Decade of Action for Road Safety 2011–2020. Geneva: World Health Organization, 2010.