STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

ADA Notice

For individuals with sensory disabilities, this document is available in alternate formats. For alternate format information, contact the Forms Management Unit at (916) 445-1233, TTY 711, or write to Records and Forms Management, 1120 N Street, MS-89, Sacramento, CA 95814.

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA16-2807		

4. TITLE AND SUBTITLE	5. REPORT DATE
OCICATS (Online Community Input Classification to Advance Transportation Services) - A GIS-based Decision-Support Tool	09/02/2016
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR	8. PERFORMING ORGANIZATION REPORT NO.
Lourdes V. Abellera	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
Civil Engineering Department California State Polytechnic University, Pomona 3801 West Temple Avenue Pomona, CA 91768	3762
	11. CONTRACT OR GRANT NUMBER 65A0529 TO 028

12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
University of California Center on Economic Competitiveness in Transportation (UCCONNECT), UC Berkeley 2150 Shattuck Avenue, Suite 300 Berkeley, CA 94704-5940	Final Report, May 1, 2015 through April 1, 2016
	14. SPONSORING AGENCY CODE Caltrans, DRISI

15. SUPPLEMENTARY NOTES

This project developed a software based on machine learning techniques to summarize discussions regarding transportation in California on social media. This tool is intended to reveal factors important to transportation users that may not be evident to transit agencies. The software uses topic modeling to cluster public messages related to transportation on the Twitter social media platform.

16. ABSTRACT

Public transportation agencies can obtain large amounts of information regarding timeliness, efficiency, cleanliness, ridership, and other performance measures. However, these metrics are based on the interests of these agencies and do not necessarily represent the concerns of the customers. Recently, social media have become a platform for people to show their satisfaction or discontent about particular services and products (e.g., Twitter feeds, Yelp reviews, Change.org petitions). This tool was intended to reveal features of ridership that are not evident to transit agencies. For instance, a sense of community and pride are positive aspects of ridership that are not measured by traditional surveys. Specifically, sentiment analysis techniques were utilized to classify numerous sets of rider sentiment data over a period of time and for particular locations (e.g., Metrolink station). Each aspect was shown as a theme in a geographic information system (GIS) layer. This online GIS-based tool can be accessed by transportation planners to determine areas of service where they can focus their resources either for the short term or long term. Project deliverables include an analysis of online social media communications to identify statements regarding public transit ridership; discovery of factors affecting ridership from social media text communications; software documentation including a report describing in detail the developed methodology, results of the software evaluation of correctness, and a summary of novel ridership factors discovered by the system. This research will be featured as an article in ACCESS Magazine.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
OCICATS (Online Community Input Classification to Advance Transportation Services) - A GIS-based Decision-Support Tool, social media, GIS-based tool, transportation planners, public transit, ridership, UCCONNECT, California transportation, Fastrack	The readers can freely refer to and distribute this report. If there are any questions, please contact one of the authors.

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
No security issues	26	Free for E-copy

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in alternate formats. For information, call (916) 654-8899, TTY 711, or write to California Department of Transportation, Division of Research, Innovation and System Information, MS-83, P.O. Box 942873, Sacramento, CA 94273-0001.

Analyzing Spread of Influence in Social Networks for Transportation Applications

Final Report
UCCONNECT 2015-2016 - TO 029

Lourdes V. Abellera, PhD
Assistant Professor

California State Polytechnic University, Pomona

Anand Panangadan, PhD
Assistant Professor
California State University, Fullerton

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The California Department of Transportation (Caltrans) provided financial support for this project through the UCCONNECT (University of California Center on Economic Competitiveness in Transportation) Faculty Research Grant for FY 2015-16.

We are grateful for the assistance of Madonna Camel, the UC CONNECT Program Manager, and Christine Azevedo, Caltrans Contract Manager.

The following are the main student contributors to this work:
Pritesh Pimpale, California State University, Fullerton
Nanwarin Chantarutai, California State Polytechnic University, Pomona
Diana Lin, California State Polytechnic University, Pomona

ABSTRACT

Using Twitter data, we developed a tool for generating a list of potential influential individuals and/or organizations for particular transportation-related topics by counting the number of mentions of a specific Twitter user and retweets of a particular tweet. Their locations are indicated in Google Maps. Although papers in the current literature propose different measures of influence using both contrived and real data, mentions and retweets are the most reliable measures of influence. To date, our work is the only work in the study of influence that is transportation-related. We believe our tool will advance the state of the practice. We have listed many purposes that can be addressed by using our tool including limiting misinformation and encouraging acceptance of a new transportation product or service.

1. INTRODUCTION

1.1 Problem Statement

In Southern California, public transit is an unpopular mode of transportation. There are many reasons for this sentiment. Communities and business districts are so spread out that one person might live in Los Angeles and work in Claremont. It is generally impractical to ride the train or the buses in this situation, unless, for example, the MetroLink can be a convenient alternative. People generally believe that public transportation is slow, unreliable, unsafe, and unclean. Some of these beliefs have an actual basis, and some do not. For example, reliability depends on a particular route, and safety depends on the area surrounding a station. Most people who do not usually ride public transportation will never be able to experience its benefit if they hear negative experiences and views of other riders. Consider the following situation. An unpleasant experience happens to a rider, for instance, if she missed her plane due to the Los Angeles Airport FlyAway shuttle being late. The disgruntled passenger, especially if young, might tell her story in Twitter or Facebook, or write about the unpleasant experience in her blog. Consequently, her followers will take note of this experience, and will affect their decisions to take or not take the LAX FlyAway shuttle for their next trip to the airport depending on the amount of influence the dissatisfied passenger has exerted on them.

In this project, we have analyzed the concept of influence in social media, in particular, the Twitter social media site, and identified the individuals who exert the most influence to those they interact with using two measures of influence, the Twitter mention and retweet. There are several studies that use social media to assess public perception and sentiment regarding public transit. Bertrand et al. (2013), using geospatial tools, discovered through Twitter that public sentiment is generally positive in public parks and negative at transportation hubs. Collins et al. (2013) use sentiment analysis methods on Twitter messages to identify specific problems (e.g., fires) on public transportation in real-time.

1.2 Relevance

This project aims to enhance the public's perception of transportation services in the hope of increasing the number of people using shuttles, buses, and trains and others who use other sustainable transportation alternatives such as biking and walking. If more people participate in sustainable transportation, then there will be fewer cars that contribute to the production of greenhouse gases (GHGs). Currently, the study of the spread of influence in social media is in theoretical stage, and there are many researchers that have offered good solutions to maximizing the influence of an entity to an event, activity, or idea (Kempe et al. 2003, Budak et al. 2011). However, all of these are still in theory and need experiments to discover an actual benefit from the procedure. In addition, this study is multidisciplinary; it has utilized concepts in sociology, psychology, computer science, as well as geospatial technology. We also argue that this project will improve the quality

of life of Californians. It is known that being a passenger in a vehicle is less stressful than driving.

Utilizing spread of influence as a method to advertise the benefits of public transit, carpooling, biking, and walking challenges the status quo because the method is extremely different from the traditional ways of disseminating information. Recently, for example, MetroLink has painted its trains to promote safety in crossing rail tracks. Of course, advertising about how you can save money by buying Tap cards instead of paying cash, or asking your employer to pay for your bus or train ride, can be seen ubiquitously not only as posters in both Foothill and Metro buses, trains, and stations, but also in brochures.

1.3 Social Media and Influence

Facebook, Twitter, YouTube, Instagram, LinkedIn, and ResearchGate are popular examples of social media platforms. In social media, users share aspects of their personal life, pass on or report news and other information, voice their opinions on a current topic, and ask for or offer an advice about a product or service. A user can also scout for a potential employee or request for an academic paper. Participants of social media form virtual communities and networks of people with similar interests and goals. The Pew Research Center has conducted a study of social media usage from 2005 to 2015 (Perrin 2015) resulting in numerous statistics that show social media's ubiquity. That is why social media is an immense venue for entities (individuals, groups, large organizations) to exert their influence, whether intentional or not, upon other users. In fact, there are services to compute for a person's influence such as Klout (URL in References) and PeerIndex (URL in References).

Merriam-Webster defines influence as "the power to change or affect someone or something". Keller and Berry (2003) identified the characteristics of 10% of the Americans who tell the rest what to buy, which political figures to vote for, and where to travel. "Word of mouth" is that potent conduit that influential people use to create change in another person's behavior. "Mouth" is the person who says the "word", with the person more important than the message. With social media platforms, "word of mouth" is even more important now as "mouth" or the user has numerous venues to express his/her "word" or content in the form of tweets, blogs, videos, and images.

1.4 Measures of Influence in Twitter

Twitter (URL in References) is one of the most prominent micro-blogging services available. Here, users can send and read "tweets" which are short messages with a maximum limit of 140 characters. Every day, about 500 million tweets are sent (Twitter Usage Statistics 2016). Advertisers of new products, presidential campaigners, and the like want to reach Twitter audience to push their content and influence the beliefs and actions of users. As such, various researchers have investigated the concept and spread of influence in Twitter using actual data (Leavitt et al. 2009) or models (Bakshy et al. 2011).

Many researchers ask how to quantify influence. Twitter collects data from user and one of the common ones is the “follower”. Twitter applies a social-networking model known as "following". Here, a user can select who he/she wants to "follow" to receive tweets from. The consent of the user to be followed is not required. It is easy to think that a user with a million followers is very influential. However, Cha et al. (2010) have shown that indegree, the number of people who follow a user, is not proof of influence. Weng et al. (2010) found that users follow those users who follow them back out of politeness and not influence; this reciprocity is called homophily. This phenomenon has also been observed by Aral et al. (2009). Retweets, the number of times others “forward” a user’s tweet, has also been considered a measure of influence. Another parameter is mentions, which is the number of times others mention a user’s name. Anger and Kittl (2011) also listed other proposed measures of influence such as Follower/Following Ratio, Retweet and Mention Ratio, and Interactor Ratio which is the number of individual users who retweet content or mention user X divided by the total number of followers of user X. Only mention and retweet prove to be consistent measures of influence (Cha et al. 2010, Leavitt et al. 2009). Hence, we used these two parameters to define our influential individuals, groups, or other entities in developing our tool. Retweet is associated with the value of the content, while mention is related to the importance of the user’s name (Anger and Kittl 2011).

2. METHODOLOGY

We initially applied the methodology we originally proposed (i.e., topic modeling, stochastic cascade modeling). We attempted to find the source A of a transportation-related tweet and how that tweet is propagating through the social network by finding the user B that retweeted the original tweet of A to users C and D. However, after the third level, it became difficult to determine how the tweet is propagating because some relevant information is not revealed by Twitter due to privacy issues. The propagation of a message in a social network has been studied to maximize the spread of influence (Kempe et al. 2003, Even-Dar and Shapira 2007, Chen et al. 2009) and to limit the spread of misinformation (Budak et al. 2011); however, these efforts are still in the theoretical stage.

As discussed in the Introduction section, the definition and measurement of influence has been studied and applied to Twitter datasets. None of these involved a transportation-related topic. Again, we define influence as mentions and retweets, and below is how we counted these parameters in our code and deployed the web application.

The code is divided in 3 parts.
1. Twitter stream capture. - Python and MongoDB
2. Java REST API. - Java
3. Web app front end - HTML and JavaScript (AngularJS)

We have used Python (URL in References) for capturing the real time tweets from the Twitter API (Application Programming Interface) (URL in References) with the help of a python library called Tweepy (URL in References). An API is a defined way for a program to accomplish a task, usually by retrieving or modifying data. It is easy to use python for accessing the Twitter API with the help of tweepy library as well as it is efficient in maintaining continuous http connection with the server while consuming minimum resources. MongoDB (URL in References) is used for the Data storage of the tweets as the data obtained from Twitter is in the JSON format which can be easily stored in the NoSQL database like MongoDB. It provides Document oriented data structure which helps in directly storing the twitter data without any change in it. It also provides the drivers for the python and java which we have used for our application.

Java (URL in References) is used as the Server side REST API for the computation of the Influential twitter handles. Java provides rich data structures like HashMap which we have used for computing the influential twitter handles using the number of re-tweets and mentions. JAX-RS (URL in References) is used for developing the REST API for the web app in Java. AngularJS (URL in References) is chosen as the client side script for the client web app. It provides a rich framework for developing single page application with communication with server using http calls. It provides an easy interfacing with the UI (User Interface) framework like Angular Material (URL in References) which is used for creating the UI of the web app.

With regards to visualization using geographic information systems (GIS), we initially believed that ArcGIS was the best platform to show the results. However, we realized ArcGIS might be unavailable for smaller transit agencies, hence, we decided to use Google Maps for this purpose. By using the free Google Maps API, we embedded the Google Maps site into our web application. The locations of the influential individual or organization and who they have influenced are shown in Google Maps. Google Maps is popular; many people are familiar with its symbols (e.g., Google Map pin, map view, satellite view, terrain view) and functions (e.g., zoom in, zoom out). Google Maps is very widely used; more than 1,000,000 web sites (Hoetmer 2013) use the Google Maps API.

The code that counts mentions and retweets and which includes the web application designed to interact with users are submitted to Caltrans in a zip file. They are also available in other formats if needed (e.g., in GitHub, URL in References). We have made initial efforts to apply for a patent for these products.

3. RESULTS AND DISCUSSION

3.1 Strengths and Limitations

One of the strengths of the code we developed is that it gathers tweets free of charge. We attempted to buy Twitter data from Gnip (URL in References) but it was too expensive. For example, for the duration from 1/1/2015 to 3/15/2016, this retrieved approximately 35,000 tweets over the 435-day period. The cost of this data license is $7,625. Because our proposed tool is being offered to transit agencies and not businesses, we cannot recommend an expensive source of data. The Twitter API harvests data only from the last seven days. However, with the code we developed, it is continuously harvesting tweets from the date it was deployed (Jun 9, 2016). The code has only a few lines and it can easily be modified. The standard APIs are all free and instructions can be easily accessed online. Our method did not reinvent the wheel. Our method made a connection between the Twitter API and Google Maps API, which are standard, accessible APIs that are already proven successes. Our methodology is also able to produce results relevant to the chosen topic with high accuracy. The next section will give an example with a detailed discussion.

One of the limitations of the tool is that we do not have a comprehensive list of transportation topics. Table 1 is a list of the keywords used to capture the live tweets from Twitter in the current version of the code. The three topics currently in the web app are just a subset of the topics we assigned. Transportation planners may be looking for very specific topics such as “Metro annual budget 2016” or “Caltrans 710 houses”. In this case, we can just easily add these new keywords in the code. Our tool will harvest the tweets for these additional keywords starting from the date those keywords were added.

Table 1: Keywords used to capture the live tweets from Twitter

#caltrans	California trains	#expresslanes California
caltrans	California freight	fastrak California
California transportation	high speed rail California	transponder California
California traffic	#hsr California	#bicyclelanes California
California cars	bullet train California	bicycle lanes California
California rail	expresslanes California	

The second limitation is that tweets with no user-specified location are not included in the results. This will limit the number of tweets processed. The third limitation is that it shows results from users that have provided their location outside of California. It is possible that (1) the user has moved recently to California and has not updated his/her Twitter account, or (2) the user does not live in California, but tweeting about a topic in California. There is no way to determine which of the two situations is correct. This situation can be handled in different ways. First, the transportation planner using the web app can disregard the tweet. As the locations of the Twitter users are shown in Google Maps, it is easy to see the user's location. However, we can also argue that we can use this to influence California residents. For example, New York City is a public transit-friendly city, and those who tweet from New York may see the California High Speed Rail positively and tweet accordingly. As Southern California is spread out, many residents are averse to taking public transit. However, tweets from the New Yorker may influence them to take public transit. The last limitation is that it may be difficult to actually find these influential individuals if we choose to contact them. Twitter has control over privacy so the actual names and addresses of these influential people will be difficult to obtain. However, for organizations like non-profits or newspapers, this is not an issue.

3.2 An Example- California High Speed Rail

Tweets are accumulating in the web app. Hence, the web app is dynamic, and not static unlike other studies. Hence, the results are a snapshot of the retweets and mentions at a particular time. To see how the web app works, we take the topic "High speed rail" on July 26, 2016. Figure 1 presents the screenshot of the web app showing the three available topics that one can choose from. These topics can be modified or removed, and more topics can be added. (The Google Maps pins, the inverted-drop-shaped icons that mark locations in Google Maps, show the locations of a previous search.) Note that this screenshot is just a portion of the complete result of the search.

Next, we click "High speed rail", then click "Mentions", then click "FIND INFLUENCERS". Figure 2 shows the results of these steps. We observe that the Washington Examiner has the most number of mentions (i.e., 19 mentions). The Washington Examiner is an American political journalism website and weekly magazine

based in Washington, D.C. If we click Washington Examiner, we see Figure 3 which shows the tweets mentioning the tweet “California's bullet train on the track to extinction” of Washington Examiner. Aside from the mention, we can see that some tweeters also add their own sentiment.

Now, if we click “Retweets”, then click “FIND INFLUENCERS”, Figure 4 is what we will see. Scott Walker has the most number of retweets at 68, followed by DownsizeTheFeds (21 tweets) which ties with Paul Rogers also at 21 tweets. (The program lists the first 10 entities.) If we click Scott Walker, we find only one tweet, in Figure 5, that was retweeted 68 times (i.e., This is why I called the "high speed rail" line between Milwaukee and Madison a boondoggle). It is interesting to note that the Google Maps pin points to Wisconsin. Indeed, Scott Walker is the governor of Wisconsin and his tweet is about a recent news item about the California high speed rail. Table 2 shows the complete result of finding influencers for mentions, and Table 3 for retweets. In these two tables, we observe that there are more retweets than mentions suggesting that content is more important than the name of the user in determining influence. We can argue that 2 to 3 mentions or 6 to 9 retweets are hardly indicative of influence. Our tool lists the top 10 potential influencers; this number can be easily modified in the code.

It is interesting to note that Scott Walker, the governor of Wisconsin, is more influential (with 68 retweets) than Jerry Brown, the governor of California (with 3 mentions). It is reasonable to assume that the California governor is much more familiar with California issues than the Wisconsin governor. However, the Wisconsin governor is probably more well-known for the topic of bullet trains. It is possible that Jerry Brown will be more influential for the keyword “traffic”, for example.

Finally, we observe that our tool finds relevant tweets with high accuracy with the current list of keywords. The user CA Water 4 All hardly qualifies as an interested entity in high speed rail. However, upon examining the individual tweets, we find that this user laments about California projects, which according to his/her opinion, are a waste of money. An example tweet is “#California’s biggest boondoggles: The Delta tunnel projects and high-speed rail”. Another interesting user is the Google Play Music. His/her inclusion in the resulting list is not an error. An example tweet is “I bet it'll be 2020 B4 @GooglePlayMusic has a sleep timer and alarm clock feature... California will have bullet trains 1st at this pace!”. This tweet is not actually about the California high speed rail, but it shows the popularity of this topic. The location of this particular tweet is Newcastle, Australia. The only irrelevant tweet included in the results is one tweet by Paul Rogers with the tweet “California imposing $45 fee on rail cars carrying toxic chemicals to fund emergency response”.

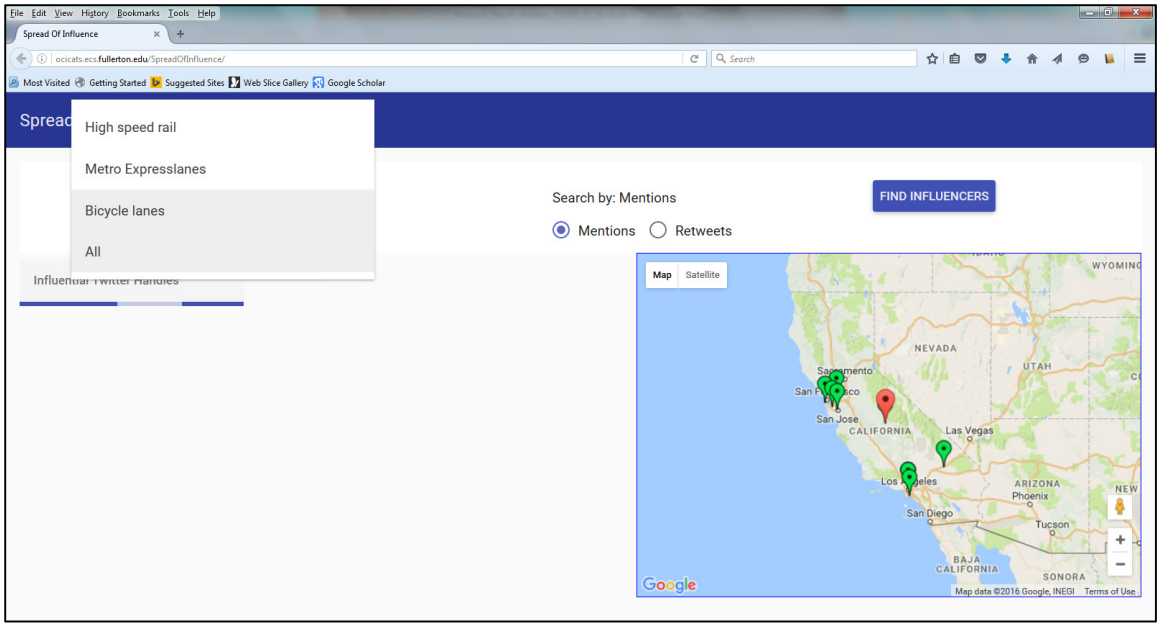

Figure 1: Available topics in the web app

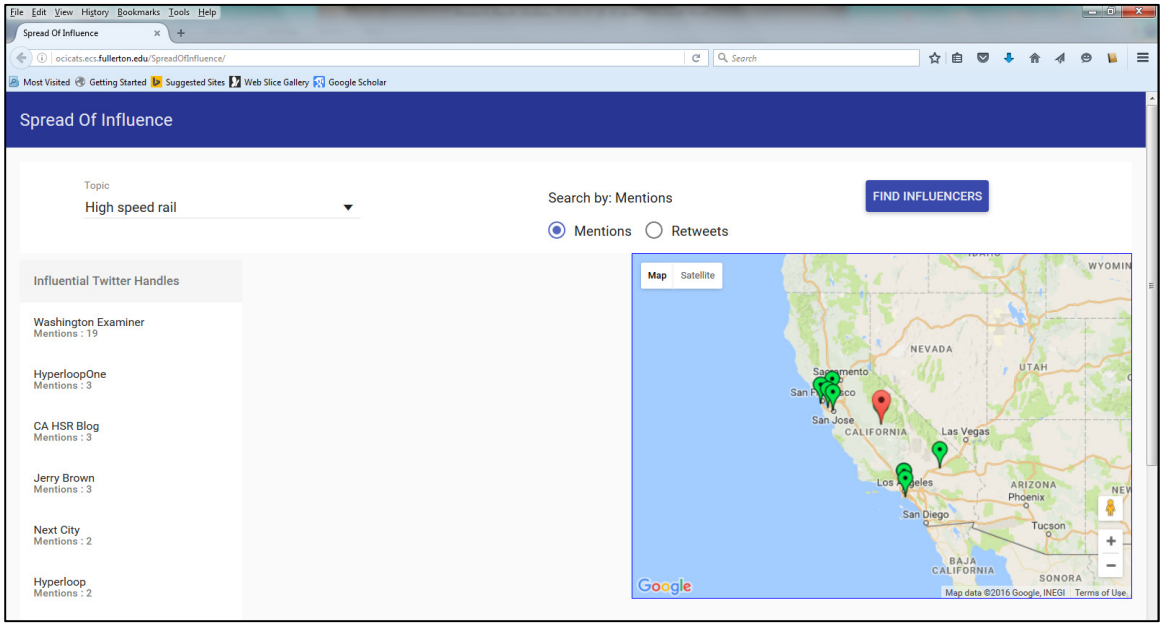

Figure 2: Result of searching for mentions for the topic "High speed rail"

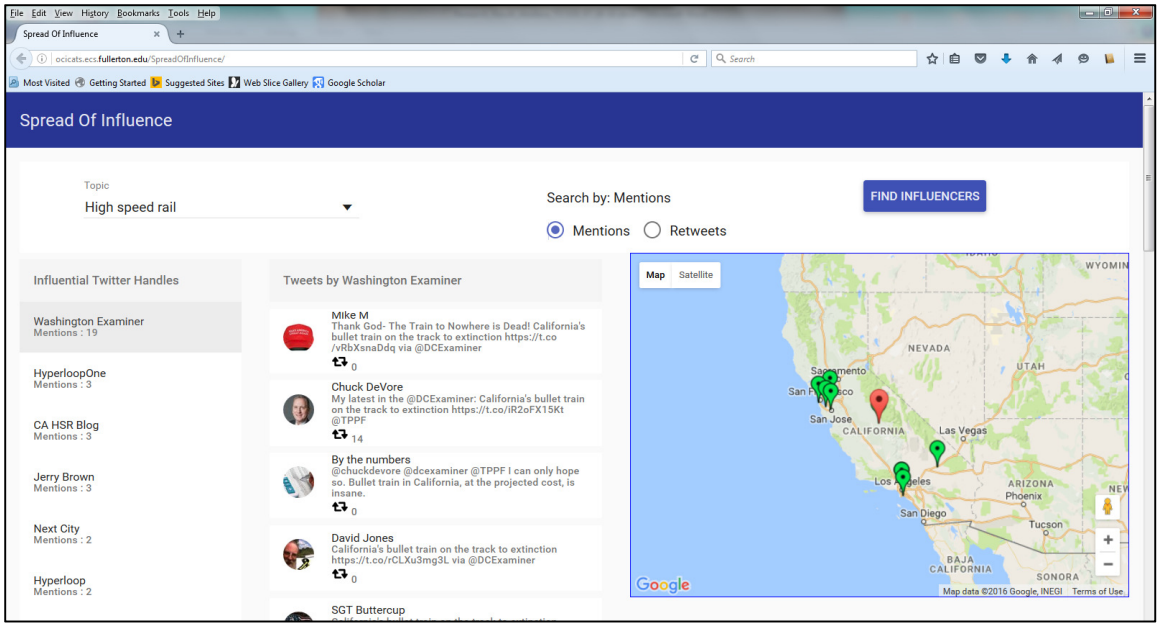

Figure 3: The tweets mentioning the tweets of Washington Examiner, the entity with the highest number of mentions

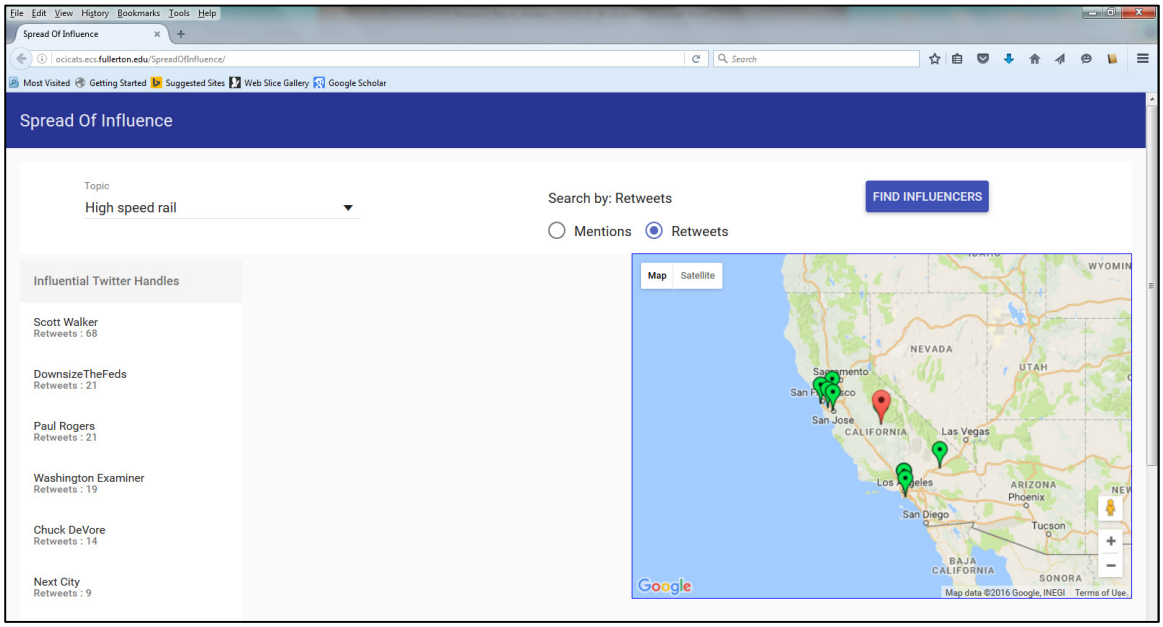

Figure 4: Result of searching for retweets for the topic "High speed rail"

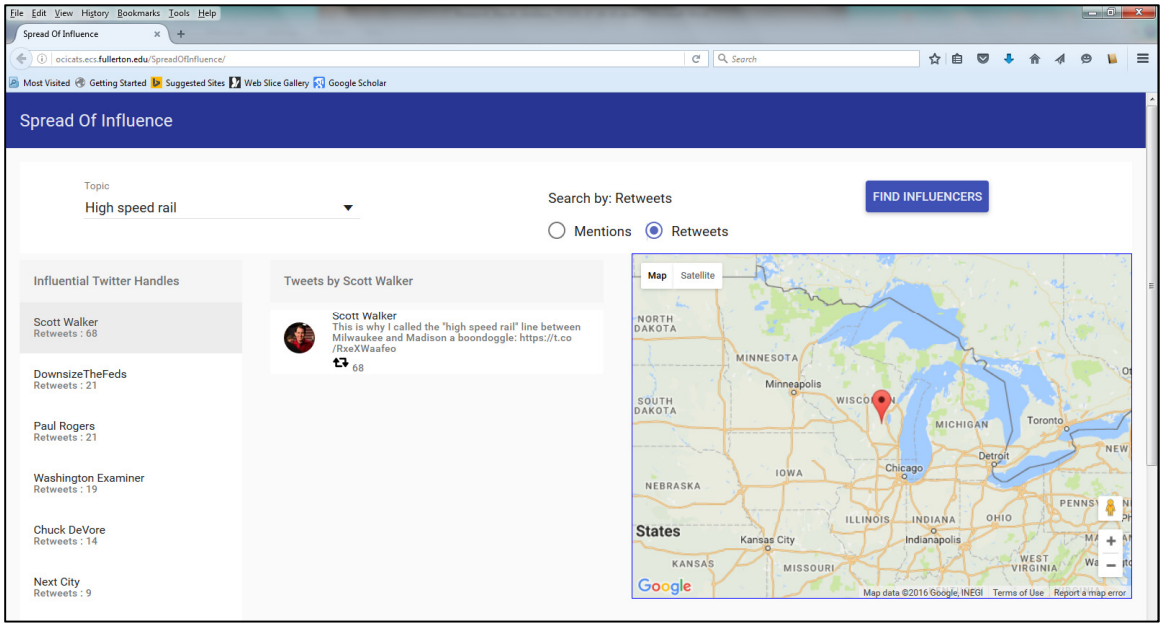

Figure 5: The tweet by Scott Walker that was retweeted 68 times

Table 2: Search results of using mention as the measure of influence

User	Number of Mentions
Washington Examiner	19
HyperloopOne	3
CA HSR Blog	3
Jerry Brown	3
Next City	2
Hyperloop	2
CA Water 4 All	2
Texas Public Policy	2
Kira Fucker @home	2
Google Play Music	2

Table 3: Search results of using retweet as the measure of influence

User	Number of Retweets
Scott Walker	68
DownsizeTheFeds	21
Paul Rogers	21
Washington Examiner	19
Chuck DeVore	14
Next City	9
Christopher Balding	9
janewells	9
Dagwood Bumstead	7
TED	6

3.3 Recommendations and Policy Implications

The following are the potential uses of the tool we developed. We hope that this tool will make an impact on the transportation profession and practice.

Limiting the spread of misinformation

When electronic toll collection (ETC) was just beginning for the and I-110 Metro ExpressLanes, there were incorrect beliefs about the system, for example, that one always has to pay a toll fee to use these highways (Caltrans and Metro, 2013). Exacerbating this was the large number of people who were fined incorrectly (Grover and Drechsler 2013); they vented their frustration in Yelp (URL in References). One probable reason is that the educational campaigns on the toll lanes were not sufficient. Many more freeways (e.g., I-405) are planned to have ETC (Caltrans GIS Data, 2016). Our tool can gage if the public information department of Caltrans is providing sufficient information to drivers.

Identifying and using celebrities

Wikipedia is rife with articles about celebrities. One type of section included in these articles is about philanthropy, activism, charity, and other causes. Many movements are already using celebrities to influence beliefs, behaviors, policies, and laws. For example, Matt Damon speaks against the Yulin Dog Meat Festival in China (Oppenheim 2016). Anne Hathaway is an outspoken supporter for lesbian, gay, bisexual and transgender (LGBT) rights (Wong 2012). With regards to sustainable transportation, Rachel McAdams moves around Toronto by bicycle and does not own a car (Schwartz 2013). Our tool can be used to identify these celebrities (e.g., Elon Musk) who advocate sustainable transportation and how much influence they exert. In fact, advocacy and expertise are not even requirements to be influential. Cha et al. (2010) proved that most influential Twitter users can exert substantial influence over various subjects, not only for the topic they are mostly known.

Changing negative publicity

A few years ago, the San Francisco-Oakland Bay Bridge received a lot of negative publicity, mostly generated by the news articles published by the Sacramento Bee (Siders 2012). Our tool can be used to know if this negative publicity has subsided. In addition, we can identify those individuals who criticize public transit using limited experience. A sample of a tweet is "I wish the Blue Line wasn't so sketchy and shady. I'd travel out more". We can approach this user and offer him/her incentives so that he/she will have a variety of public transportation experiences, and hopefully will change his/her beliefs. Examples of incentives are free rides and gift cards.

Knowing the initial perception of a proposed action

The gas tax is declining because vehicles have become more fuel-efficient and therefore people buy less gas. Funding road maintenance and repair has become a challenge. The road charge is being examined as an alternative for generating revenue. Several states have conducted pilot tests on road charges including Nevada, Washington, Minnesota, and Oregon; California will begin its own pilot program soon. There are other options to increase funds for road maintenance. Increasing vehicle license fees, gas tax, and sales taxes are suggestions. It is also possible to collect tolls in more highways or divert money from other areas like education and health care (California Road Charge 2016). Our tool can be used to assess the perceptions of Californians to the road charge, how much they are willing to pay, and their concerns about privacy if the determination of the vehicle miles traveled (VMT) will involve collecting information about their location.

Assessing the success of social media efforts of Caltrans

Caltrans has lots of social media efforts. For example, #IwillRide is a hashtag that promote the acceptance, support, and future use of the high speed rail. However, a quick Google search on July 28, 2016 resulted in only about 1,500 results. Clearly, that is not the amount of publicity Caltrans needs for a multi- billion-dollar project. Our tool can be used to identify how the #IwillRide hashtag is being propagated in Twitter and determine if the sentiments of these tweets are positive or negative.

Promoting a good image of Caltrans and the transit agencies

If Caltrans and the transit agencies are viewed positively by the public, they will likely support their objectives one of which is to generate revenue for maintaining transportation-related assets. Some income-generating projects may require the vote of the electorate. For example, MTA hopes that taxpayers will approve a $120-billion sales tax increase (Nelson 2016). To dramatize the importance of trust in a public agency, in July 2012, there was a referendum on a 1-cent transportation special-purpose local-option sales tax (T-SPLOST) to pay for traffic and road improvements in Metro Atlanta. Because of the voters' distrust of the governing bodies and lack of cooperation among themselves, these voters rejected the $7.2 billion transportation plan (Hart 2012). Our tool can be used to identify both the negative and positive influencers.

Catching road culprits

Freeway paint bombing, the splattering of paint on roads, has been observed. Caltrans has an effort to catch the culprits by giving instructions for law enforcement (Hoover 2015).

Our tool can be used to see how many people have been reached by this appeal to catch criminals.

3.4 Other Work

Throughout the duration of this project, many other related tasks were performed. We developed several versions of a Twitter crawler (i.e., a tool for collecting the relevant contents of a set of tweets). We also looked at other social media platforms, for example, Instagram, to manually and automatically determine the influential people and organizations talking about positive, neutral, or negative things about Caltrans through the #Caltrans tag. In the duration of the public survey for the development of the 2018 California State Rail Plan between January 27 and March 4, 2016, we gathered over 5,000 tweets in the same time period and annotated the sentiment manually. The official Caltrans survey received 2,189 responses (2018 California Rail Plan Survey Summary Report, 2016).

Many of the efforts leading to the goals of this project have been presented and the details are shown in Appendix 1. The supporting materials are available upon request.

4. CONCLUSIONS AND FUTURE WORK

Using Twitter data, we developed a tool for generating a list of potential influential individuals and/or organizations for particular transportation-related topics by counting the number of mentions of a specific Twitter user and retweets of a particular tweet. Their locations are indicated in Google Maps. We believe our tool will advance the state of the practice. We have listed many purposes that can be addressed by using our tool including limiting misinformation and encouraging acceptance of a new transportation product or service. Below we discuss how we can improve this tool and expand its use.

The list shown in Table 1 includes keywords related to "Bicycle lanes" and "Metro Expresslanes", but we are not getting related tweets as these keywords might not be the exact keywords users are using on Twitter for these topics. Most of the tweets we are able to capture are related to traffic and high speed rail. More words should be incorporated in the code, and test if the search are resulting in correct tweets.

We can ask practicing transportation professionals (e.g., planners, engineers) to suggest and/or decide on their own keywords based on an actual issue they have. We can recommend that they perform a Google search first to see what terms people write, know, or hear about. For example, in the city of Walnut, there is a service, called Dial-A-Cab for the elderly and disabled residents of the city. However, this service is underutilized. It is hoped that social media will expedite the promotion and use of this program (Personal communication with Karen Villasenor, City of Walnut employee, January 2016.) A quick search on Twitter Advanced Search (URL in References) on July 28, 2016 generated zero results for "Dial-A-Cab Walnut". Future work can study how topics with a more limited audience can be included in our tool.

Another future work can be automatically identifying the keywords to be used in Twitter search for a new transportation-related topic. For example, #iwillride is hardly related to the California high speed rail. An automated process that will discover these keywords from Twitter communications will be an interesting topic in the future.

The transit agencies have started using social media to promote sustainable transportation. The OCTA (Orange County Transportation Authority) produced a video (OCTA Video 2016) in an effort to encourage people to ride the OCTA buses. It was done professionally; it was trying to be cool and funny. However, users have noticed the unrealistic portrayals in the short video. One might ask why some videos become much more popular than others which are called "viral" videos. Weng et al. (2012) investigated this phenomenon with Twitter datasets and concluded that users' limited attention is a strong factor in determining whether a tweet would become viral. Therefore, future work can be a study of how a tweet, Instagram photo, or YouTube video becomes viral for a target audience and how this can change behavior.

5. REFERENCES

2018 California Rail Plan Survey Summary Report, 2016

Anger, I., & Kittl, C. (2011, September). Measuring influence on Twitter. In Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies (p. 31). ACM.

AngularJS. https://angularjs.org/

Angular Material. https://material.angularjs.org/latest/

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544-21549.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74). ACM.

Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., & Bar-Yam, Y. (2013). Sentiment in new york city: A high resolution spatial and temporal view. *arXiv preprint arXiv:1308.5010*.

Budak, C., Agrawal, D., & El Abbadi, A. (2011, March). Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web* (pp. 665-674). ACM.

California Road Charge, 2016. http://www.californiaroadchargepilot.com/about/faqs/.

Caltrans and Metro (2013). 110 ExpressLanes Performance Update, as of 2/28/13.

Caltrans GIS Data (2016) California Express Lanes (2015), Shapefile

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM, 10*(10-17), 30.

Chen, W., Wang, Y., & Yang, S. (2009, June). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 199-208). ACM.

Collins, C., Hasan, S., & Ukkusuri, S. V. (2013). A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation*, 16(2), 2.

Even-Dar, E., & Shapira, A. (2007, December). A note on maximizing the spread of influence in social networks. In *International Workshop on Web and Internet Economics* (pp. 281-286). Springer Berlin Heidelberg.

GitHub. https://github.com/

Gnip. https://gnip.com/

Grover, J. and Drechsler, P. (2013). Thousands of Incorrect Violation Notices Sent to SoCal Drivers. *NBC Los Angeles*.

Hart, A. (2012). Voters reject transportation tax. *Atlanta Journal-Constitution*.

Hoetmer, K. (2013). A fresh new look for the Maps API, for all one million sites. *Geo Developers Blog*.

Hoover, K.L. (2015). Freeway paint bombing proves pernicious new pollution. *Mad River Union*.

Java. https://java.com/en/download/

JAX-RS. https://jax-rs-spec.java.net/

Keller, E., & Berry, J. (2003). The influentials: One American in ten tells the other nine how to vote, where to eat, and what to buy. Simon and Schuster.
Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137-146). ACM.
Klout. https://klout.com/
Leavitt, A., Burchard, E., Fisher, D., & Gilbert, S. (2009). The influentials: New approaches for analyzing influence on twitter. *Web Ecology Project*, 4(2), 1-18.
MongoDB. https://www.mongodb.com/
Nelson, L.J. (2016). MTA shows what taxpayers will get if they OK a $120-billion sales tax increase. *Los Angeles Times*.
OCTA Video (2016) Dump the Pump with OCTA on OC Bus. https://www.youtube.com/watch?v=0Ri_Z8bMl8E
Oppenheim, M. (2016). Yulin Dog Meat Festival: Matt Damon, Pamela Anderson and Leona Lewis call for controversial event to be banned. *The Independent*.
PeerIndex. https://www.brandwatch.com/peerindex-and-brandwatch
Perrin, A. (2015). Social media usage: 2005-2015.
Python. https://www.python.org/
Schwartz, J.D. (2013). Rachel McAdams Gets Around Toronto By Bike. *The Urban Country*.
Siders, D. (2012). Caltrans asks for retraction of Bee story on Bay Bridge flaws. *The Sacramento Bee*.
Tweepy. http://www.tweepy.org/
Twitter. https://twitter.com/
Twitter Advanced Search. https://twitter.com/search-advanced?lang=en
Twitter API. https://dev.twitter.com/overview/api
Twitter Usage Statistics, 2016. http://www.internetlivestats.com/twitter-statistics/
Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010, February). Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 261-270). ACM.
Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific reports, 2*.
Wong, C. (2012). Anne Hathaway To Donate Money From Wedding Photos To Gay Marriage Advocacy Groups. *The Huffington Post*.
Yelp. http://www.yelp.com/biz/metro-expresslanes-gardena

APPENDIX: Poster Presentations of Related Work

Chantarutai, N., Lin, D., Hao, L., Panangadan, A., and Abellera, L.V., "Identifying Sources of Negative Tweets on Transportation Services," UCCONNECT Student Conference, Feb. 11-12, 2016 (Riverside, California, U.S.A.).

Abstract:
Social networks are quickly growing into the greatest influencers in people's lives. The speed and ease with which information propagates through social media allows both facts and personal opinions to spread rapidly. At times however, the information and opinions that are spread along social networks may be misleading or misrepresentative of individuals or organizations, and can lead to the spread of negative influence. Such dispersion of negative sentiments may be detrimental to the image of service providers. In particular, since there is generally a negative disposition towards transportation hubs, this can be especially damaging to public transportation service providers. In order to lessen the scope of this spread of unfavorable perceptions, it is important for service providers to be able to identify both the source of negative sentiments as well as the most influential players involved in its spread through social networks.

Villasenor, K., Chan, H., Hao, L., Panangadan, A., and Abellera, L.V., "An Analysis of the Interactive Dialogues Created through the Caltrans Hashtag on Social Media Platforms," UCCONNECT Student Conference, Feb. 11-12, 2016 (Riverside, California, U.S.A.).

Abstract:
As public transportation companies grapple with the challenge of increasing ridership, company officials seek to improve opinions of their public transportation services through social media platforms. As a result, we are seeing public transportation companies utilizing social media platforms to promote their reliability, safety, and various other positive characteristics of their services. As social media platforms allow communications to be turned into interactive dialogues, information disseminated by public transportation companies through social media platforms is of great importance; however, there is also a great opportunity found in public transportation posts made by individuals who are not associated with public transportation companies.
This study examines how individuals not associated with public transportation companies are non-purposefully using Instagram to improve the reputation of the public transportation industry. Specifically, we examine the Caltrans hashtag on Instagram. We reviewed a sample of Instagram posts with the Caltrans hashtag and coded for a variety of factors including numbers of likes, positive or negative posts and comments, numbers of followers, and frequency of irrelevant posts with the hashtag #caltrans. Together, non-associated individuals are omnipotent in terms of directly and indirectly impacting Caltrans's reputation due to their large social networks. With effective interaction between public transportation companies and non-associated individuals, we expect that the public transportation industry as a whole may be viewed in a significantly more positive manner.