JUSTICE POLICY CENTER

Public Perceptions of Police on Social Media

A Big-Data Approach to Understanding Public Sentiment toward the Police

Ashlin Oglesby-Neal, Emily Tiry, and KiDeuk Kim

February 2019

The millions of tweets shared on Twitter daily are a rich resource of public sentiment on countless topics. In the wake of highly publicized officer-involved shootings, many people take to social media to express their opinions, both positive and negative, of the police. We collected millions of public tweets to explore whether we can measure public sentiment toward the police. Specifically, we examine how public sentiment changed over time and in response to one high-profile event, the death of Freddie Gray in Baltimore on April 19, 2015 after suffering from a spinal injury while in police custody. While accounting for the larger trends in the public image of the police on Twitter, we find that sentiment became significantly more negative after Gray's death and during the subsequent protests.

Important to community policing is the ability of law enforcement and community members to form partnerships and identify problems and solutions together (Goldstein 1987; Trojanowicz and Bucqueroux 1990). The widespread use of social media makes it a potential resource for identifying issues in police-community relations. Law enforcement agencies are increasingly interested in using social media to learn from and engage with the public, with an eye toward enhancing police-community relations. However, there is little guidance, let alone evidence-based research, to help the police understand the potential of social media. This project explores the feasibility and utility of Twitter data for measuring public sentiment toward police.

It is likely that public sentiment of police changes over time, and in response to prominent events. These events bring police actions to the forefront of the public's attention, and may lead them to express their opinions through tweets. Although one may suspect that criticisms would become more frequent after events involving alleged police misconduct, people may also want to voice their encouragement or support of the police. In this analysis, we examine both positive and negative tweets about the police over time.

Data

To examine sentiment toward police, we acquired more than 65 million public tweets that include the words "police" or "cop" from a Twitter data provider. We use Twitter data rather than other types of social media data – more specifically, Facebook – for a few reasons. First and foremost, Twitter is easy to acquire and use. It is public by default and provides a platform that facilitates a discussion around big issues of the day. Second, Facebook offers a lot more privacy controls than Twitter (e.g., control who can connect with you, control who can see your photos, control whether users can message you). As a result, Facebook is not as facilitative as Twitter in terms of sharing information freely and publicly, which is an important aspect of how the public forms an opinion of the police.

The project team therefore acquired data from Twitter, covering January 12 through June 12 of 2014 and 2015, which provides 304 days of public tweets about law enforcement—approximately 150 days before and after the death of Freddie Gray in Baltimore in 2015 and the same period in 2014 as a control to account for any seasonal trends in public sentiment.

We examine these intervals because examining all relevant tweets over an extended period beyond 2014 and 2015 would be computationally cumbersome. Twitter data include not only the actual text of the tweet, but also information about the user and tweet. This other information includes date and time; the user's number of followers, friends, and tweets; and whether the account is verified. We used both the text of the tweets and their metadata to develop a model that classified the sentiment of each tweet.

Method

Sentiment Prediction

Before examining whether public sentiment changed over time or in response to high-profile events, we needed to determine the sentiment of each tweet. The sheer number of tweets precluded any individual from reading and classifying all tweets. Instead, we manually classified a sample of 4,050 tweets into four categories: positive, negative, neutral, and not applicable. Then, we used these tweets to build a model that could predict the sentiment of the remaining tweets.

TABLE 1

Sentiment Categories and Example Tweets

Sentiment	Example tweets
Positive	▪ "I totally respect cops even tho when they pull me over I second guess lol" ▪ "Nick from police and safety is my hero #freedom"
Negative	▪ "I hate the majority of all cops" ▪ "The amount of police brutality has gotten way out of hand."
Neutral	▪ "Police Search for Suspect After Man Shot Multiple Times in South Sacramento" ▪ "Police: Federal judge shot during robbery attempt in Detroit"
Not applicable	▪ "so I just saw a commercial for mall cop 2 and I'm pretty hype" ▪ "Think Imma Cop This Iphone 6 When It Drop"

To build a sentiment prediction model, we extracted as much information as possible from each tweet. Before extracting this information, we cleaned the tweets to remove mentions, hashtag symbols, URLs, punctuation, and stop words (e.g., and, the, to). We then transformed the text of the tweet into a document-term-matrix, where every word became a variable. We also used Stanford University's natural language processing tool, CoreNLP,[1] to identify named entities in the tweet (i.e., location, organization, person, date, or time) and parts of speech so we could indicate if the word "cop" was being used as a verb. We counted the number of positive and negative words in each tweet according to Hu and Liu's sentiment word list (2004). We also created an indicator for whether police were being referenced as a source of information, which is frequently done by news organizations.

Besides information in the text of the tweet, we also used Twitter metadata to predict the outcome (i.e., sentiment). Examples of this include the user's account age, number of followers, and number of tweets, along with whether the account is verified. We also made variables to indicate whether the tweet included a URL, mention, emoji, or hashtag. In addition, we considered the day of the week and the time of day the tweet was posted. In total, we used more than one hundred features to develop the sentiment prediction model. The 16 most important variables in the final model, as determined by the machine learning algorithm, are listed below.[2] Individual words are denoted in quotation marks.

1. URL indicator
2. Any location mention
3. Total negative words
4. "Police"
5. Any named entity mention
6. Length of tweet
7. User statuses count
8. "Cop"
9. Police source indicator
10. Cop used as verb
11. Age of user's account
12. "Black"
13. "Call"
14. Total positive words
15. "Fuck"
16. "Brutal"

We tested several machine learning algorithms, and the top-performing model used gradient boosted regression for classification, with an overall accuracy of 63 percent. We used this model to predict the sentiment of all the tweets. The most common sentiment was neutral (58.9 percent). The rates of not applicable (19.6 percent) and negative (19.3 percent) were very similar. Positive tweets were uncommon and composed 2.0 percent of tweets about police.

Compared with extant research on public sentiment that typically deals with a binary outcome (e.g., positive or negative), the performance of our prediction model is less than desirable. However, this is largely because our model predicts one of four discrete values, and that is computationally more complicated and challenging. Our analysis did not focus just on positives versus negatives because that dichotomy does not reflect the complex reality of social media data.

TABLE 2
Sentiment of Tweets about Police

Sentiment	Frequency	Percentage of all tweets
Positive	1,323,142	2.0
Negative	13,344,742	19.4
Neutral	38,247,598	58.9
Not applicable	11,859,964	19.6
Total	**64,775,446**	**100.0**

Source: Authors' analysis of Twitter data for January 12–June 12, 2014, and January 12–June 12, 2015.
Note: Table shows sentiment of tweets that mention "police" or "cop" over the 2014 and 2015 periods mentioned above.

Sentiment Analysis

To analyze the sentiment of tweets about police, we compared daily tweet totals and rates before and after Freddie Gray's death on April 19, 2015, to the same days in 2014. In effect, we constructed a quasi-experimental evaluation based on a difference-in-differences framework (Ashenfelter and Card 1985). The framework essentially calculates differences between the pre (a few months before April 19) and post (a few months after April 19) time frames for both the treatment (2015) and control (2014) groups, then subtracts the differences between the two groups. The outcome is the daily tweet totals or rates of each sentiment (e.g., the daily rate of negative tweets or the daily total of positive tweets).

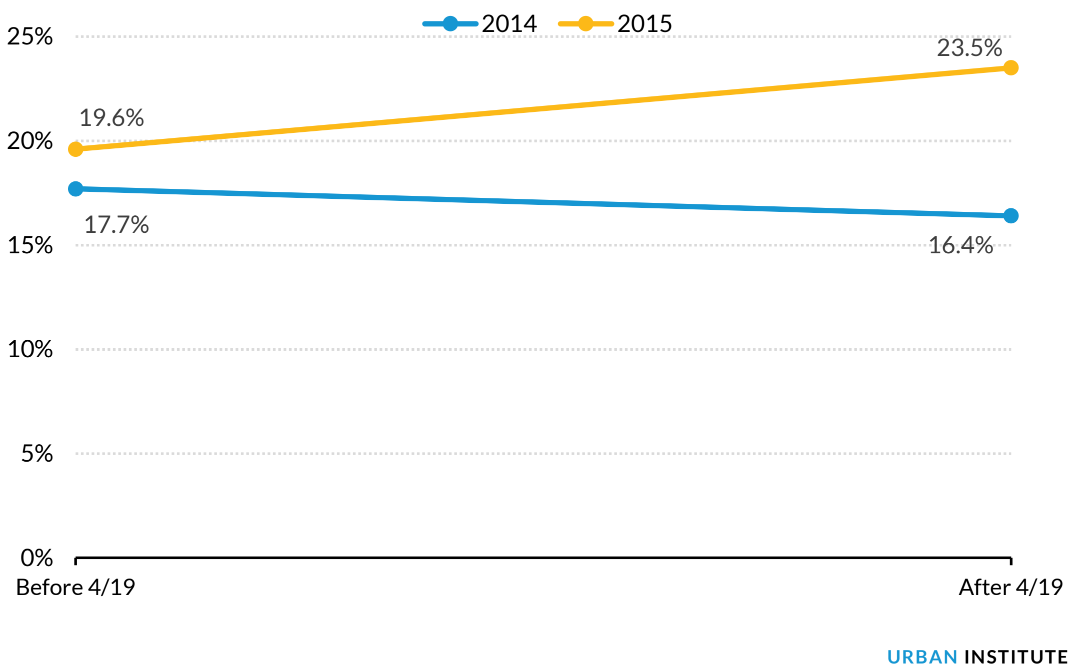
Results

Public sentiment toward police changed over time and mainly became more negative. Positive sentiment toward police was relatively stable in 2014 and 2015, and the rate of positive tweets in both years was about 2 percent. In contrast, the rate of negative tweets was 17 percent in 2014 and 21 percent in 2015, showing an increase in negative sentiment over time. Tables A.1–A.4 in the appendix provide summary statistics on the negative and positive tweets by time frame.

Looking within years, sentiment fluctuated more in January through June of 2015 than during those same months in 2014. Most of the 2015 fluctuations were in negative sentiment. In the three months before Freddie Gray's death on April 19, 20 percent of tweets about police were negative. In the two months after his death, 23 percent of tweets about police were negative (figure 1). The total count of negative tweets increased after his death, spiking during the protests in Baltimore the following weeks. The total count of positive tweets also increased, as more people were tweeting about the police, but the share of positive tweets remained the same.

FIGURE 1

Average Daily Percentage of Tweets about Police that are Negative



Source: Authors' analysis of Twitter data for January 12–June 12, 2014, and January 12–June 12, 2015.

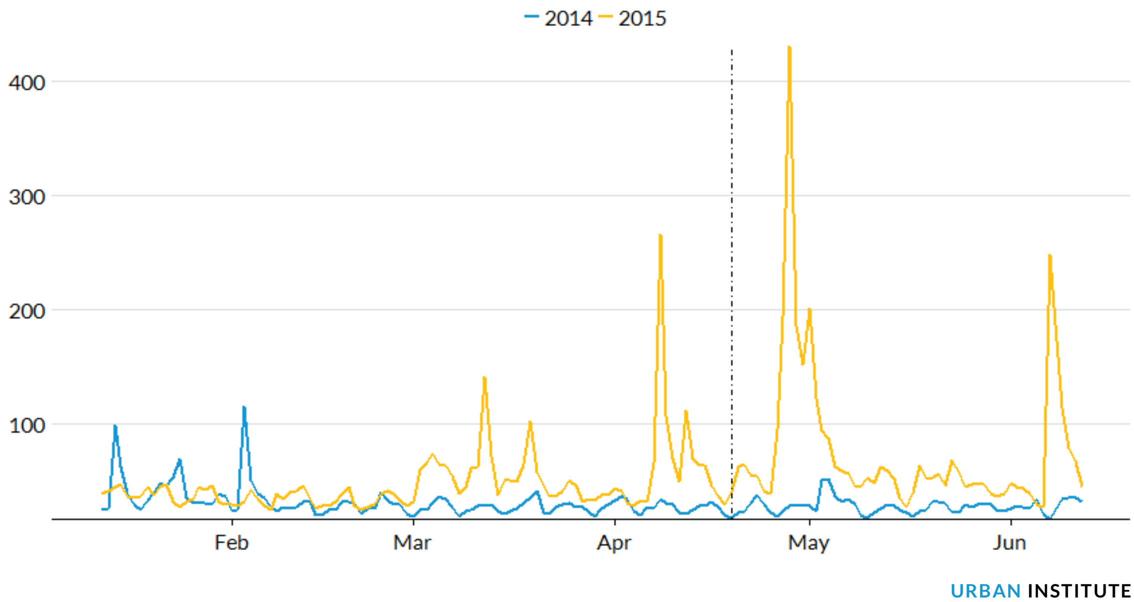
Figure 2 shows the daily negative tweet frequencies. It has several clear peaks, one in the beginning of April, likely in response to the death of Walter Scott on April 4, 2015, and another in late April through the beginning of May 2015, likely related to the ongoing protests in Baltimore and Freddie Gray's funeral on April 27. The negative tweets then return to their baseline levels and peak again during June 7–9, 2015, possibly in response to the incident at a pool party in McKinney, Texas, on June 5, where an officer restrained a teenage girl. The frequency of negative tweeting relates to several high-profile police incidents in 2015.

After examining the fluctuations in sentiment over time, we estimated the effect of Freddie Gray's April 19, 2015, death on public attitudes toward police using a difference-in-difference design. We find that compared with the same days in 2014, the daily rate of negative tweets increased 5 percentage points after Gray's death. This increase is statistically significant at the 0.01 level. There was no effect on the rate of positive tweets (table 3).

FIGURE 2

Total Daily Negative Tweets about Police

In thousands



Source: Authors' analysis of Twitter data for January 12–June 12, 2014, and January 12–June 12, 2015.
Note: Dotted vertical line marks the day of Freddie Gray's death in 2015.

TABLE 3

Difference-in-Difference of Positive and Negative Sentiment Rates

Standard errors in parentheses

	Dependent Variable	
	Rate positive	**Rate negative**
Year	−0.0001	0.019***
	(0.0005)	(0.007)
After April 19	−0.636	−100.377***
	(1.528)	(23.473)
Difference-in-difference	0.0003	0.050***
	(0.001)	(0.012)
Constant	0.188	−37.857***
	(0.911)	(13.991)
Observations	304	304
R^2	0.004	0.180
Residual standard error (df = 300)	0.003	0.049
F Statistic (df = 3; 300)	0.374	21.882***

Source: Authors' analysis of Twitter data for January 12–June 12, 2014, and January 12–June 12, 2015.
**p < 0.05; ***p < 0.01

We also examined the effect of Gray's death on the total number of positive and negative tweets. There were statistically significant increases in the daily frequencies of both positive and negative tweets about the police after Gray's death compared with the same days in 2014 (table 4). The average daily positive tweets increased by 1,579, while the negative tweets increased by 33,922. The increase in negative tweets was much greater than the increase in positive tweets.

TABLE 4

Difference-in-Difference of Positive and Negative Sentiment Tweet Frequencies

Standard errors in parentheses

	Dependent Variable	
	Total positive	**Total negative**
Year	1,140***	15,787***
	(215)	(4,946)
After April 19	−3,180,217***	−68,321,302***
	(728,003)	(16,717,545)
Difference-in-difference	1,579***	33,922***
	(361)	(8,299)
Constant	−2,292,869***	−31,763,792***
	(433,919)	(9,964,318)
Observations	304	304
R^2	0.31	0.20
Residual Std. Error (df = 300)	1,508	34,624
F Statistic (df = 3; 300)	45***	26***

Source: Authors' analysis of Twitter data for January 12–June 12, 2014, and January 12–June 12, 2015.

$^{**}p < 0.05$; $^{***}p < 0.01$

Implications

The key finding of this project is that public sentiment towards the police became significantly more negative after Gray's death and during the subsequent protests. This expected finding should be no surprise to anyone. However, it speaks to the validity of our approach to assessing public sentiment through Twitter data, and we were able to quantify the magnitude of change in public sentiment over time.

In other words, this analysis of Twitter data demonstrates that public sentiment toward police can be measured using publicly available social media data and that these data can be a resource for law enforcement agencies. Although we examined the effect of an extremely high-profile event on public attitudes, individual agencies can modify this method to fit their community relation needs. For example, a police department could collect the tweets mentioning the department's Twitter handle and classify them by sentiment. The agency then could monitor public sentiment in real time and use that feedback to develop or target community engagement activities. Some guidance on this process is available in the Social Media Guidebook for Law Enforcement Agencies. A police department also could partner with a research organization or social media analytics company to measure sentiment. Social media data are not a full measure of a community's opinions of police, but they can be another tool to support community policing efforts.

Although Twitter provides a promising source of public sentiment toward police, sentiment analysis using tweets has some limitations. Tweets are short and often express immediate reactions in incomplete sentences, making it difficult to determine an overarching message or sentiment. Misclassification of the sentiment of a tweet happens, even with advanced modeling and natural language processing techniques. Further, people often use sarcasm in tweets, which can be difficult to detect and lead to more misclassification. Despite the difficulties in accurately determining the sentiment of tweets, our analysis reveals that sentiment can be classified fairly reliably.

This analysis also reveals how a police incident in one community can affect attitudes toward police across the country. We measured the effect of Freddie Gray on sentiment toward police using tweets from the entire United States, not just the Baltimore area. The statistically significant effect on negative sentiment demonstrates how high-profile events can shape public discourse at large. The instantaneity and accessibility of social media may help drive the ability of public sentiment to change quickly.

Even though measuring public sentiment through tweets can be a starting point for efforts to build police-community relations, this measure is not enough to fully gauge and understand public sentiment. Tweets only capture the universe of people who use Twitter, and what they can briefly share in a few sentences. To better understand public opinion, agencies could host town hall meetings to discuss the causes of positive or negative sentiment. Further, a robust understanding of the public's opinions toward police would likely necessitate surveying all types of community members and soliciting feedback on a host of topics related to the police. Because such surveying is costly and time-consuming, police departments could consider innovative methods, like examining tweets, to understand how the public perceives their activities.

Appendix. Summary Statistics

TABLE A.1

Daily Negative Tweet Totals by Date

Date	Mean	Standard deviation	Minimum	Maximum
2014				
Before April 19	31,185	14,016	17,812	115,383
After April 19	27,788	6,742	16,610	51,294
2015				
Before April 19	46,972	29,474	24,211	265,882
After April 19	77,496	69,216	26,661	429,988

TABLE A.2

Daily Negative Tweet Rates by Date

Date	Mean	Standard deviation	Minimum	Maximum
2014				
Before April 19	0.177	0.039	0.116	0.413
After April 19	0.164	0.031	0.107	0.279
2015				
Before April 19	0.196	0.050	0.128	0.439
After April 19	0.233	0.072	0.150	0.497

TABLE A.3

Daily Positive Tweet Totals by Date

Date	Mean	Standard deviation	Minimum	Maximum
2014				
Before April 19	3,502	756	2,207	6,274
After April 19	3,502	685	2,083	5,812
2015				
Before April 19	4,642	1251	2,813	10,644
After April 19	6,222	2,914	3,146	16,402

TABLE A.4

Daily Positive Tweet Rates by Date

Date	Mean	Standard deviation	Minimum	Maximum
2014				
Before April 19	0.020	0.003	0.014	0.031
After April 19	0.021	0.003	0.015	0.028
2015				
Before April 19	0.020	0.003	0.013	0.029
After April 19	0.021	0.004	0.015	0.043

Notes

[1] For more information on CoreNLP, see https://stanfordnlp.github.io/CoreNLP/.

[2] In gradient boosting machine (gbm) models, each variable has a level of importance from 0 to 100. Variables with greater predictive importance have higher values.

References

Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs." *Review of Economics and Statistics* 67:648–60.

Goldstein, Herman. 1987. "Toward Community-Oriented Policing: Potential, Basic Requirements, and Threshold Questions." *Crime and Delinquency* 33 (1): 6–30.

Hu, Minging, and Bing Liu. 2004. "Mining and Summarizing Customer Reviews." In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–77. New York: Association for Computing Machinery (ACM).

Trojanowicz, Robert C., and Bonnie Bucqueroux. 1990. *Community Policing: A Contemporary Perspective.* 1st ed. Cincinnati, OH: Anderson Publishing Company.

About the Authors

Ashlin Oglesby-Neal is a research associate in the Justice Policy Center at the Urban Institute. Her research includes developing and validating risk assessment tools and evaluating the impact of sex offender treatment on recidivism. Her analytic capabilities include data processing and management, regression, survival analysis, text mining, machine learning, and data visualization.

Emily Tiry is a research associate in the Justice Policy Center, where her current research focuses on the development and validation of risk assessment tools and the effect of health care on reentry outcomes. She has extensive experience with a variety of quantitative research methods, including evaluation methods, predictive modeling, and machine learning.

KiDeuk Kim is a senior fellow in the Justice Policy Center, where he leads multidisciplinary research teams to examine issues related to criminal justice interventions and policies. He was a principal investigator on the social media project for which this report was developed. His current research focuses on understanding decisionmaking in the criminal justice system and evaluating criminal justice programs.

Acknowledgments

This project was supported, in whole or in part, by cooperative agreement number 2015CKWXK016 awarded by the U.S. Department of Justice, Office of Community Oriented Policing Services. The opinions contained herein are those of the author(s) or contributor(s) and do not necessarily represent the official position or policies of the U.S. Department of Justice. References to specific individuals, agencies, companies, products, or services should not be considered an endorsement by the author(s) or contributor(s) or the U.S. Department of Justice. Rather, the references are illustrations to supplement discussion of the issues.

The internet references cited in this publication were valid as of the date of publication. Given that URLs and websites are in constant flux, neither the author(s) or contributor(s) nor the COPS Office can vouch for their current validity.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

Thank you to Graham MacDonald, Alex Engler, Kyle Ueyama, and Aaron Williams of the Urban Institute for their contributions to the sentiment classification process.

URBAN INSTITUTE

2100 M Street NW
Washington, DC 20037

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.