The Impact of Social Media Trends on Motor Vehicle Collisions

Ryan Longmuir

March 18, 2019

This paper investigates the causal effect of viral social media events on motor vehicle collisions. Leveraging individual Twitter posts and county level crash data, I construct a balanced panel of viral social media events. I then employ multivariate and fixed effects regression models and find an increase in social media activity is not associated with in increase in the number of motor vehicle collisions that occur during a viral social media event. These results are consistent with the conclusions of similar papers, however, I speculate the results are biased by omitted factors and measurement error in social media activity.

Introduction

Motor vehicle collisions are an issue of national public safety. The National Highway Traffic Safety Administration (NHTSA) reported over 7,277,000 motor vehicle collisions in 2016, in which 37,461 people were killed and an estimated 3,144,000 people were injured (NHTSA, 2016). Similarly, the Centers for Disease Control and Prevention (CDC) reported motor vehicle collisions as the leading cause of death of individuals between the ages of 5 and 24 in the United States (CDC, 2018). Local governments have recognized this epidemic as a call to action and have deployed interventions to combat motor vehicle collisions. For Instance, New York City mayor Bill de Blasio established Vision Zero in 2014, an initiative to eliminate all traffic deaths in New York City by 2024. Thus far programs such as Vision Zero have successfully improved street safety, however, there is still an imperative to study the potential causes of motor vehicle collisions. Investigating the causes of motor vehicle collisions provides local governments valuable insights necessary to write informed legislation.

In this paper I study the impact of viral social media events on motor vehicle collisions and attempt to answer the following questions (1) Do viral social media events cause an increase in the number of motor vehicle collisions? (2) If so, does the presence of other determinants of motor vehicle collisions exaggerate this effect? I use individual Twitter posts and county level crash data to construct a balanced panel of forty five viral social media events. Then, I use multivariate and fixed effect regression models to isolate the effect of social media use on motor vehicle collisions. The results suggest an increase in social media activity is not associated with a higher number of motor vehicle collisions, having a statistically insignificant effect of zero in all model specifications. These results are consistent with conclusions of similar papers, however, I speculate these results are biased by omitted factors and measurement error in social media activity.

There are many papers that study the effect of driver cellphone use on motor vehicle collisions, however, much of this research lacks relevance in the present time. A majority of the papers that estimate the effect of driver cellphone use on motor vehicle collisions were conducted before the popularizing of social media and the rapid increase in smart phone adoption that began in the early 2010s. Since 2011, the proportion of U.S. adults who own a smartphone has increased by more than forty percent and cellphones have evolved to boast a broader range of capabilities (Pew, 2018). The advent of smart phones, social media applications and push notifications have changed the way users interact with their devices, introducing new avenues for drivers to engage in distracted driving. This study differentiates itself from the previous body of literature as it utilizes a novel data source, Twitter posts, to study the effect of viral social media events on motor vehicle collisions.

Background

This research is similar to a large body of literature that measures the effect of driver cellphone use on motor vehicle collisions. This previous research has used a variety of methodological strategies and found conflicting evidence regarding the extent to which driver cellphone use contributes to motor vehicle collisions.

Redelmeier and Tibshirani investigate crash records and phone bills for 699 Toronto drivers recently involved in a minor car collision (Redelmeier, 1997). Redelmeier and Tibshirani estimate the effect of using a cellphone while driving on the risk of a motor vehicle collision by comparing mobile phone call-logs for drivers in non-serious accidents just before the event and at a comparable time in the past. They estimate an increase in the relative likelihood of a crash by a factor of 4.3 and no statistical difference between handheld and hands-free devices.

In 2003 Prieger and Hahn conducted a survey of 7,000 drivers to study cellphone use and driver patterns (Hahn, 2003). In contrast to Redelmeier and Tibshirani, Prieger and Hahn collect a sample that includes drivers who have not been involved in a motor vehicle collision and drivers who do not use cell phones. They find the effect of cell phone use on accidents varies across the population, in part because individuals with higher mobile phone usage are riskier drivers independent of their mobile phone usage. Prieger and Hahn's results suggest that previous estimates of the effect of cell phone use on motor vehicle collisions, based on accident-only samples, may be overestimated.

The National Highway Traffic Safety Administration conducted a naturalistic driver study which equipped 100 cars with data collection hardware to measure the effect of various distractions on crashes in authentic driving conditions (Dingus, 2006). The NHTSA collected approximately 2,000,000 vehicle miles and at least one year of data for each vehicle. Researchers found that drivers were 2.8 times more likely to have an accident while dialing a cellphone, and 1.3 times more likely to have an accident while talking on a cellphone, compared to having no distractions. This study considered driving conditions and found that performing secondary tasks (i.e., using a cellphone) while driving increased accident risk on dry road conditions only, not on wet road conditions.

Kolko used state-level panel data on mobile phone ownership and traffic fatalities to assess the overall effect of phone ownership and hands-free laws on fatalities (Kolko, 2007). The study used the NHTSA's Fatality Analysis Reporting System (FARS) micro-data on persons, vehicles, and conditions to create customized state-year totals for overall fatalities and fatalities under specific conditions, such as when roads are wet or at nighttime. Mobile phone ownership is associated with more traffic fatalities, but only in bad weather or wet road conditions. Mobile phone ownership contributes more to fatalities during rush hour than at other times, but the effect is not statistically significant.

Bhargava and Pathania exploit a natural experiment that arose when mobile phone providers offered "off-peak" pricing plans with unlimited calling starting at 9 pm in 2005 (Bhargava, 2007). Their difference-in-differences estimates the effect of mobile phone usage on accidents and fatalities around 9 pm is negative and not statistically significant. Bhargava and Pathania consider that the treatment effect around 9 pm could be different from the true effect, so they supplement their quasi-experimental approach with an analysis of aggregate mobile phone ownership and accident rates over time. They find no statistically significant relationship.

To summarize, the estimated effect of cellphone use on the likelihood of a motor vehicle collision ranges from no statistically significant difference in Bhargava and Pathania to 4.3 times in Redelmeier and Tibshirani. Research has also found that the effect of driver cellphone use on the number of motor vehicle collisions is dependent on road condition such as in Kolko and the NHTSA's naturalistic driver study.

Data and Research Design

Recall this paper aims to answer the following questions (1) Do viral social media events increase the number of motor vehicle collisions? (2) If so, does the presence of other determinants of motor vehicle collisions exaggerate this effect? I hypothesize that an increase in social media activity will have a positive effect on the number of motor vehicle collisions, but only in adverse road conditions.

Event Structure

Before discussing the data sources it is necessary to define a viral social media event and how time is recorded. Viral events are associated with an event (e.g., a sporting event or award show) and only include observations collected from a specific time and location. Viral social media events are four hours long and sum the number of Twitter posts published, motor vehicle collisions and other covariates over ten-minute intervals. This implies that each viral event has twenty four ten-minute intervals, each with values describing the number of posts published and car collisions that occurred. Viral social media events begin two hours before and end two hours after the interval with the highest number of posts published (hereafter, the peak). A viral event is only considered for analysis if the quantity of posts published during the peak is at least three standard deviations away from the mean number of posts published in an interval during that social media event.

This paper studies a total of sixty five viral social media events that take place in Los Angeles and San Francisco, California between 2013 and 2015. Fifty events are associated with sporting events and the remainder are associated with award shows.

Social Media Activity

As mentioned briefly in the previous section, social media activity is measured by the number of Twitter posts published during a ten-minute interval. Individual Twitter posts are collected using Crimson Hexagon, a program certified by Twitter that enables users to search and export Twitter posts given specific parameters such as time, location and content. In particular, users may specify location by country, state, and county. Users specify the content of their search by using a series of keywords specific to the subject they are studying. Examples of keyword specifications are included in the appendix.

A limitation of Crimson Hexagon is that the program imposes limits on the number of posts which users can export from an individual search. If a search returns an excess of ten thousand posts, Crimson Hexagon collects a random sample from the population of posts that satisfy the parameters. Since all forty five viral social media events are associated with searches that return more than ten thousand posts, social media activity is not measured directly and is vulnerable to measurement error. However, since the posts are randomly sampled, samples are unbiased and accurately represent the population of posts for each event.

Since Twitter's launch in 2008, it has experienced a gradual increase in monthly active users. In particular, the number of monthly active users grew twenty five percent between 2013 and 2015 (Twitter, 2018). To properly compare events across years, I regressed the number of posts published during a ten-minute interval on year and month fixed effects. By replacing the number of Twitter posts published during a ten-minute interval with the residual produced from this linear model, the variation associated with the yearly and monthly trends were removed. Figure 1 plots the number of Twitter posts published in an interval on the y-axis and time on the x-axis for three events that each take place in a different year. Figure 2 plots the residuals on the y-axis and time on the x-axis for the same three events.

Car Collisions

Car collisions are measured using the State Data System (SDS), a collection of digital state crash data files organized by the NHTSA's National Center for Statistics and Analysis (NCSA). There are currently thirty-four states participating in the program. Although the NHTSA's Fatality Analysis Reporting System (FARS) is a robust data set used frequently in previous literature, it has not been considered for this study because it only contains fatal car collisions. Since there are approximately forty thousand fatal crashes and six million total crashes reported each year in the United States, the SDS provides significantly more observations.

Model

To isolate the effect of social media activity on motor vehicle collisions, the general model below includes both day of the week and hour fixed effects as well as additional independent variables not absorbed in the fixed effects. As mentioned before in the Event Structure section, a viral social media event has twenty-four ten-minute intervals. Hereafter, an arbitrary ten-minute interval and viral social media event will be denoted as interval t and event i respectively. The dependent variable, $CRSH_{it}$, is equal to the number of motor vehicle collisions that occurred during event i and interval t. The first explanatory variable, TWT_{it}, is equal to the number of Twitter posts published during event i and interval t. The second explanatory variable, WET_{it}, is equal to the number of motor vehicle collisions that occurred during event i and interval t associated with adverse road conditions (i.e., wet, icy or muddy roads). The third explanatory variable, $DARK_{it}$, is equal to the number of motor vehicle collision during event i and interval t associated with dark road conditions (i.e., roads with no street lamps or broken street lamps). The fourth explanatory variable, ALC_{it}, is equal to the number of motor vehicle collisions the occurred during event i and interval t associated drivers under the influence of drugs or alcohol. Epsilon is a random error term assumed to comply with the full ideal conditions underlying the classical linear regression model which would result in Best Linear Unbiased Estimates of the coefficients when the model is estimated by Ordinary Least Squares (OLS). The general model is the following

$$CRSH_{it} = \beta_0 + \beta_1 \cdot TWT_{it} + \beta_1 \cdot WET_{it} + \beta_3 \cdot DARK_{it} + \beta_3 \cdot ALC_{it} + \epsilon$$

The second model specification includes TWT_{it}^2 to test whether the relationship between social media activity and car collisions is non linear. The third model specification includes $TWT_{it} \cdot WET_{it}$ and $TWT_{it} \cdot DARK_{it}$ to test whether an increase in social media activity has a greater effect on the number of motor vehicle collisions in adverse road conditions. The fourth model specification includes day and hour fixed effects.

Results

The results of each model are summarized in Table 2. Recall the purpose of this paper is to answer the following questions (1) Does driver cellphone use increase the number of motor vehicle collisions that occur during events characterized by an instantaneous increase in social media activity? (2) If so, does the presence of other determinants of motor vehicle collisions exaggerate this effect? The results suggest an increase in social media usage is not associated with a higher number of motor vehicle collisions, having a statistically insignificant effect of zero in all specifications. Furthermore, interaction terms between social media usage and road conditions were also statistically insignificant implying an increase in social media activity does not have a greater effect on the number of car collisions in adverse driving conditions. As expected, adverse road conditions have a positive and statistically significant effect on the number of collisions at the one percent level.

Discussion

The results produced in this paper have limitations. There are issues of unresolved endogeneity and measurement error, therefore, the estimated coefficients for social media activity in each of the models likely do not reflect the true effect of social media activity on motor vehicle collisions during a viral social media event.

There exists measurement error in the explanatory variable measuring social media activity. As mentioned in the data section, Crimson Hexagon randomly samples posts from searches exceeding ten thousand posts. Although the sample is chosen randomly and may be representative of the population of posts, it does not measure the population directly causing an unknown level of measurement error. Assuming the measurement error has zero mean, as measurement error increases, the estimated coefficient of the explanatory variable is biased towards zero. White noise surrounding social media activity could be an explanation for the estimated coefficients of zero produced in each of the models.

Additionally, there are variables this paper has yet to consider that effect motor vehicle collisions and covary with social media activity causing omitted variable bias. For instance, traffic congestion is likely negatively correlated with social media activity (e.g., people are less likely to drive during the super bowl or the academy awards). Since the number of car collisions is positively correlated with traffic congestion, the effect of traffic congestion is captured in the estimated coefficient for social media activity biasing it downwards. Omitted factors such as traffic congestion may also be the cause for the estimated coefficients of zero produced in each of the models.

Measurement error and omitted variable bias involving social media activity provide convincing evidence that the regression models constructed will not produce coefficients that accurately portray the true effect of viral social media activity on motor vehicle collisions. Further work must be done to acquire data sets that allow for the inclusion of omitted variables in the regression model and better methods of collecting Twitter posts to avoid measurement error.

References

Bhargava, S., Pathania V. 2007. “Driving Under the (Cellular) Influence: The Link Between Cell Phone Use and Vehicle Crashes.” *SSRN Electronic Journal* .

CDC. 2018. “10 Leading Causes of Injury Deaths by Age Group Highlighting Unintentional Injury Deaths, United States - 2017.”.

URL: *https://www.cdc.gov/injury/wisqars/LeadingCauses.html*

Dingus, T. A., Klauer S. G. Neale V. L. Petersen A. Lee S. E. Sudweeks J. . . . Knipling R. R. 2006. “The 100-Car Naturalistic Driving Study: Phase II - Results of the 100-Car Field.” *PsycEXTRA Dataset* .

Hahn, R. W., Prieger J. E. 2003. “The Impact of Driver Cell Phone Use on Accidents.” *SSRN Electronic Journal* .

Kolko, J. 2007. “The Effects of Mobile Phones and Hands-Free Laws On Traffic Fatalities.” *SSRN Electronic Journal* .

NHTSA. 2016. “Summary of motor vehicle crashes: 2016 data. (Traffic Safety Facts. Report No. DOT HS 812 580).”.

Pew, Pew Research Center. 2018. “Mobile Fact Sheet.”.

URL: *http://www.pewinternet.org/fact-sheet/mobile/*

Redelmeier, D. A., Tibshirani R. J. 1997. “Association between Cellular-Telephone Calls and Motor Vehicle Collisions.” *New England Journal of Medicine* .

Twitter. 2018. “Number of monthly active Twitter users in the United States from 1st quarter 2010 to 4th quarter 2018 (in millions).”.

URL: *https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/*

Figures and Tables

Figure 1: Posts Published

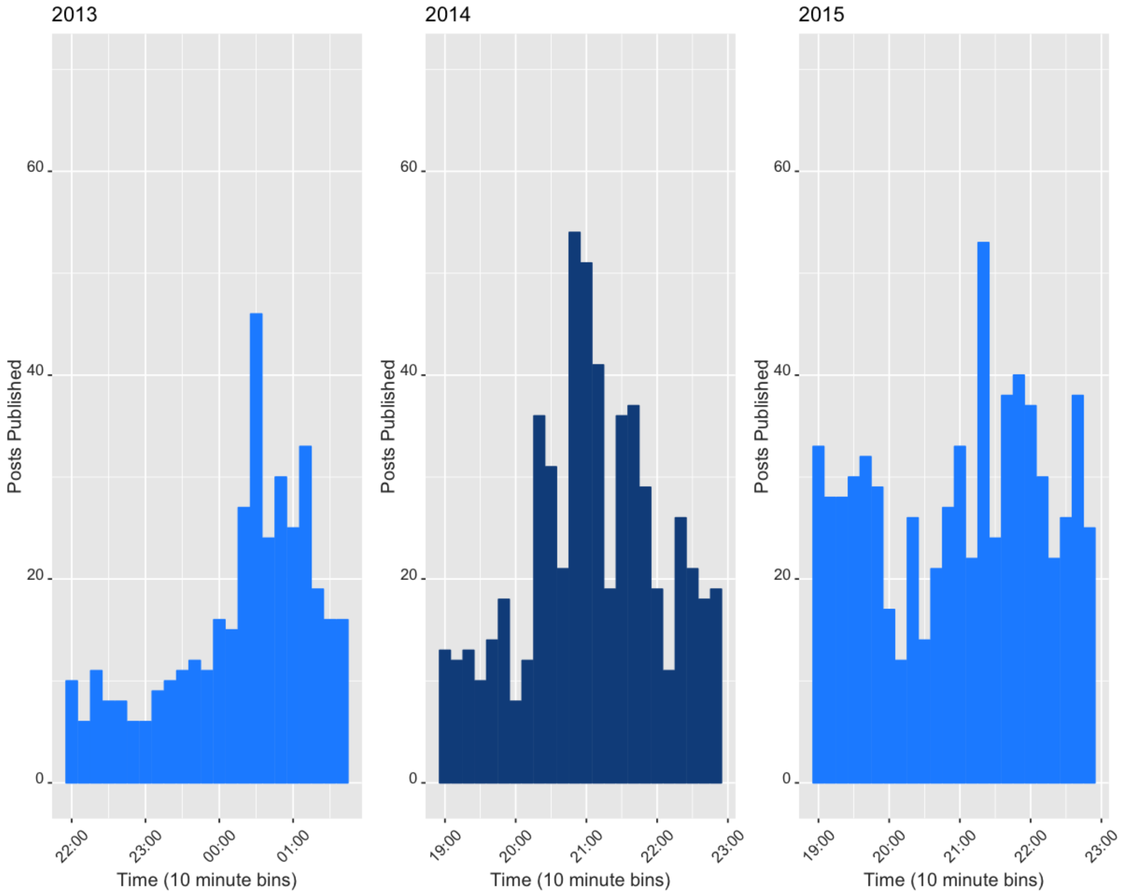

Figure 2: Posts Published Residuals

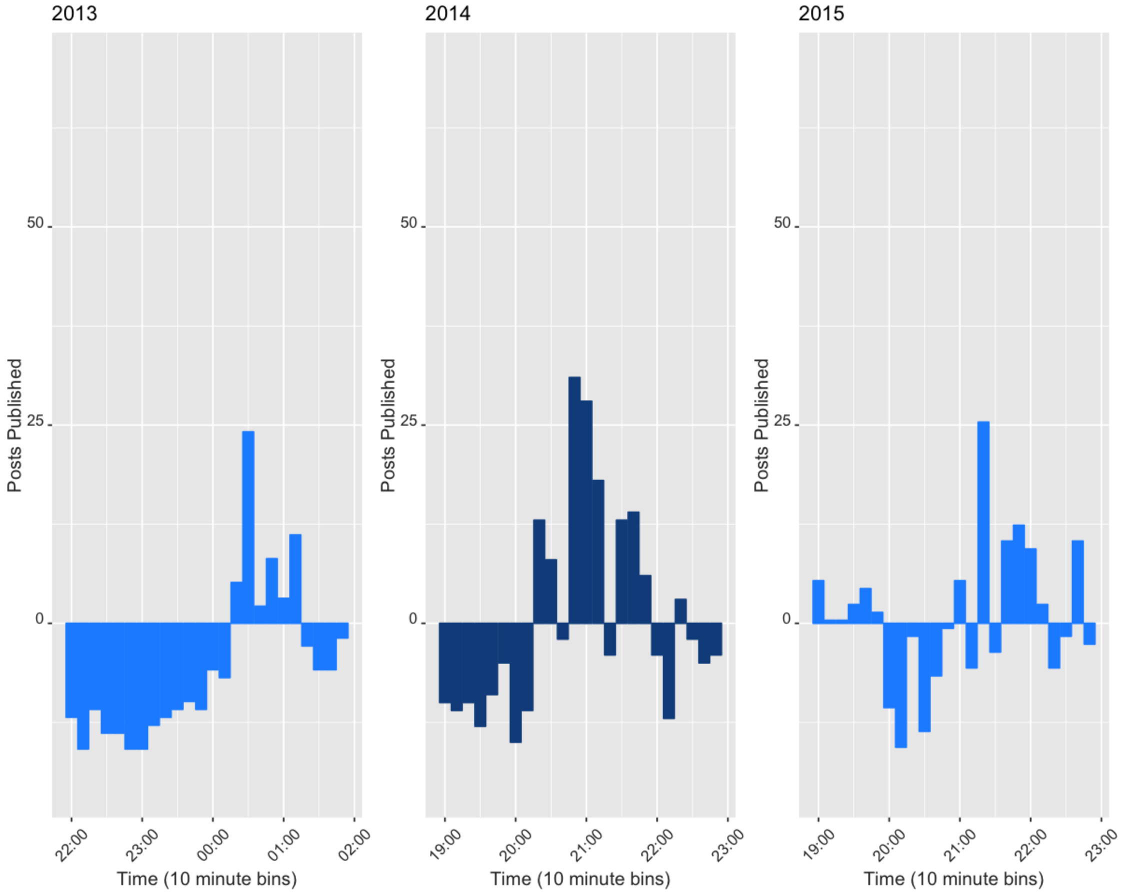

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Posts Published	1,080	24.30	17.32	0	14	30	216
Posts Published Squared	1,080	890.29	2,022.88	0	196	900	46,656
Car Collisions	1,080	0.07	0.26	0	0	0	2
Alcohol	1,080	0.002	0.04	0	0	0	1
Wet Road Conditions	1,080	0.005	0.07	0	0	0	1
Dark Road Conditions	1,080	0.01	0.12	0	0	0	1
Posts Published Residuals	1,080	−0.00	17.22	−22.95	−10.55	5.28	194.19
Posts Published Residuals Squared	1,080	296.23	1,402.17	0.0001	22.14	211.69	37,709.3

Table 2: Regression Models

	Dependent variable:			
	Number of Motor vehicle Collisions			
	(1)	(2)	(3)	(4)
Posts Published Residuals	0.0002	0.001	0.0002	−0.0000
	(0.0004)	(0.001)	(0.0004)	(0.0004)
Posts Published Residuals Squared		−0.0000		
		(0.0000)		
Wet Road Conditions	0.74***	0.74***	0.82***	0.76***
	(0.11)	(0.11)	(0.14)	(0.11)
Dark Road Conditions	0.89***	0.88***	0.87***	0.91***
	(0.06)	(0.06)	(0.06)	(0.06)
Alcohol	0.13	0.13	0.25	0.11
	(0.18)	(0.18)	(0.22)	(0.18)
(Posts Published)(Dark Road Conditions)			0.002	
			(0.004)	
(Posts Published)(Wet Road Conditions)			0.02	
			(0.02)	
Day and Hour Fixed Effects	No	No	No	Yes
Constant	0.05***	0.06***	0.05***	0.23***
	(0.01)	(0.01)	(0.01)	(0.04)
Observations	1,080	1,080	1,080	1,080
R^2	0.22	0.22	0.22	0.24
Adjusted R^2	0.21	0.21	0.21	0.22
Residual Std. Error	0.23 (df = 1075)	0.23 (df = 1074)	0.23 (df = 1073)	0.23 (df = 1057)
F Statistic	73.77*** (df = 4; 1075)	59.28*** (df = 5; 1074)	49.36*** (df = 6; 1073)	15.09*** (df = 22; 1057)

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix

Keyword Examples

The following is a key word specification used to collect Twitter posts from a televised sporting event that took place in Los Angeles during 2014.

("Lakers" OR "National Basketball Association" OR "NBA" OR "Staple Center" OR "Vander Blue" OR "Carlos Boozer" OR "Kobe Bryant" OR "Jabari Brown" OR "Jordan Clarkson" OR "Ed Davis" OR "Wayne Ellington" OR "Jordan Hill" OR "Wesley Johnson" OR "Ryan Kelly" OR "Jeremy Lin" OR "Ronnie Price" OR "Julius Randle" OR "Robert Sacre" OR "Nick Young")

Similarly, the following is a key word specification use to collect Twitter posts from a televised arard show that took place in Los Angeles during 2015.

("Emmy" OR "Comedy Series" OR "Drama Series" OR " Variety Series" OR "Miniseries" OR "Television Movie" OR "Reality-Competition Program" OR "Lead Actor" OR "Lead Actress" OR "Supporting Actor" OR "Modern Family" OR "The Big Bang Theory" OR "Orange Is the New Black" OR "Silicon Valley" OR "Veep" OR "Breaking Bad" OR "Downton Abbey" OR "Game of Thrones" OR "House of Cards" OR "Mad Men" OR "True Detective" OR "The Colbert Report" OR "Jim Parsons" OR "Julia Louis-Dreyfus" OR "Lena Dunham" OR "Melissa McCarthy" OR "Amy Poehler")