The Spark of the #MeToo Movement: Analysis of Early Twitter Conversations

Dr. Sepideh Modrek, Bozhidar Chakalov
San Francisco State University

Abstract

We characterize the public discourse in the early conversation of the #MeToo movement from Twitter data. We document and examine the content of Tweets in the first week of the #MeToo movement focusing on novel Tweets with first-person revelation of sexual assault and abuse and early life experiences of sexual abuse and assault. We use machine learning methods, Least Absolute Shrinkage and Selection Operator regressions and Support Vector Machine models, to summarize and classify the content of individual Tweets. We estimate that 11% of novel Tweets revealed details about the writer's experience of sexual abuse or assault and 5.8% revealed early life experiences of such events. These data suggest that the mass sharing of personal experiences of sexual abuse and assault contributed to the reach of this movement, and we estimate that 6 to 34 million Tweeter users may have seen such first-person revelation.

Introduction

Online communication platforms like Facebook, Twitter, and Instagram are increasingly used by public health officials (Coiera, 2013) and activists as tools to educate and amplify social movements (Freelon, McIlwain, & Clark, 2016). Yet, understanding how messages catch on and spread via social media, and the reach of hashtag movements is in its infancy. Little is known about why certain movements emerge and proliferate rapidly while others develop more slowly over time. For example, the largest online movement in recent years, #BlackLivesMatter, took over a year before it was widely used while the #MeToo movement exploded almost immediately. [1]

This brief concerns the #MeToo movement on Twitter. We document how this conversation revealed widespread and graphic sexual abuse and assault among women from all walks of life. The revelations often shared alarming early life experiences of assault and abuse adding to the more commonly told stories of sexual harassment that were already visible in the media and on Twitter at the time.

In order to better understand the content of the shared messages, we use simple machine learning tools to categorizes the content of individual-level Tweets to demonstrate the gravity of revelations and the astounding their reach.

Data

We applied for and gained access to the Twitter Premium API platform. Using the platform, we purchased both the counts and full Tweets from historical Twitter data between Oct. 14-Oct 21, 2017 which included the first week of the movement. First, we gathered the hourly counts for all Tweets that included the exact phrase MeToo (N= 1,595,453). We then limited the counts to novel English language Tweets, which exclude retweets, Tweets with images, links, or other media (N=295,021); in an effort to capture the voices of the persons who initially created the Tweet's content (Appendix 2). Finally, we concentrated only on novel English language Tweets with 'MeToo' in text and geotagged information placing the Tweet in the US (N=12,337**).**[2] The analytic sample for our content analyses comprise the full text and associated metadata for 97% of these Tweets, (See Appendix 3).

Methods

We first graphed the time patterns of all #MeToo Tweets, novel English #MeToo Tweets, novel English US based #MeToo Tweets (Appendices 4 and 5) in the first week to compare the time patterns of the different subsets of Tweets. We further show the geographic distribution of the novel geotagged US Tweets (Appendix 6).

[1] In the first 10 months of the #MeToo movement there have been approximately 1,800,000 novel English language Tweets including the phrase MeToo of which 16% originated in the first week. In contrast the #BlackLiveMatter hashtag was used for over a year before the event of Ferguson, MO when it became a focal point for the movement. See Appendix 1 for a comparison of Tweet counts #BlackLivesMatter and #MeToo in the first year of the movements.

[2] Many individuals chose not to allow Twitter to capture their location data.

We then randomly selected a subset of 650 Tweets from the novel English US based #MeToo Tweets, categorizing each on one of two two main dimensions. First, did the Tweet reveal a first-person experience of sexual abuse or assault. Revelations of sexual harassment where excluded as were brief Tweets that lacked sufficient detail to determine the nature of one's experience, such as a simple "Me Too". From among those Tweets that revealed a first-person experience of sexual abuse or assault, we further classified those where the revelations were about an experience of sexual abuse or assault in early life (age 22 or younger). We then created a word-level feature-matrix for each Tweet.[3] We used Least Absolute Shrinkage and Selection Operator (LASSO) regression models on the training sample to find most predictive words for revelation of sexual abuse and assault and revelations of early experience of sexual abuse and assault. The goal was to succinctly summarize the key words/elements posted.

We then used Support Vector Machine (SVM) models on the same training sample of 650 to predict the multidimensional words/features (based on over 11,000 words) to separate Tweets with a revelation of sexual abuse and assault from all others. We use these models to classify the remaining full sample of Tweets (N=11,285). We assess the quality of the SVM model's classification with test sets to estimate positive and negative predictive values (PPV and NPV).

Based on these classifications and empirically derived estimates of the 25-75% range of the number of followers from each poster who revealed an event, we calculate a lower bound on the reach of the Tweets in the first week of the movement. This number represents the number of potential Tweeter users who would have been exposed to these revelations.

Results

Appendix Figures 4 & 5 show the timeline of MeToo Tweets illustrating how the movement grew. We show hourly counts of Tweets from Oct 14, 2017 00:00 GMT, the day before the movement began. We compare all Tweets, to the Novel English Tweets, to the Novel Geotagged English Tweets– the main sample used in this study – to show that they follow a similar overall time trend in the first week.

We next turn to the 650 randomly selected Tweets in the training set from the analytical sample. Of these Tweets 19% were classified as having a first-person revelation of sexual abuse or assault.[4] We use this data in LASSO regressions to identify the most predictive words for sexual abuse/assault. Table 1 summarizes the most predictive words related to revelations of sexual abuse/assault. Not surprisingly the most predictive verbs included 'Rape', 'Grope', 'Grab' and 'Shout' among others. Other highly predictive words were related to intoxication ('Drunk', 'Drug') and early life experiences ('Age', 'College'). About five percent of the Tweets in the training set had phrases or words that indicated an early experience of sexual abuse or assault. Table 2 summarizes the most predictive words related to early experience of sexual abuse or

[3] Word-Level matrices create a matrix for all the words that are in all 650 training set Tweets. Then each Tweet is treated as an observation and each word is treated as a feature or word variable. If a Tweet has a particular word in it then that feature or word variable is designated as a one. It is otherwise a zero.

[4] See Appendix 3 for examples of the categorizations, which were done by one of the authors using the "Rape Culture Pyramid" created by Ranger Cervix & Jaime Chandra.

assault. While many of the predictive terms are the same, the list of persons – or possible assailants – listed for those who had an early life experience include terms like ‘Neighbor’ and ‘Step-Father’. The verb list also includes ‘Rip’ and ‘Afraid’.

We next used SVM to build a classifier to distinguish Tweets with revelation of sexual assault and abuse from all other Tweets in the remaining 11,285 Tweets in the analytical sample but not in the training set. Based on the SVM algorithm, 1,287 Tweets or 11% had a revelation of sexual abuse or assault and 657 Tweets or 5.8% had a revelation of early sexual abuse or assault (See Table 3). To assess the validity of the SVM classification against a human based classification, we calculate both the PPV and NPV. We find that the SVM classification has a PPV of 96%, meaning that 96% of the Tweets deemed as revealing sexual abuse or assaults by the SVM algorithm were confirmed by the research team. The NPV was 86% suggesting that the algorithm may slightly under report the number of Tweets with revelations of sexual abuse or assault. For early experience classification, the PPV is 81% and the NPV is 98%. In this case the false positive rate was slightly higher but the false negative rate is essentially negligible. Together this suggest that our algorithm is preforming well, and the manual classification was highly congruent with the machine generated classifications.

We then mapped the proportion of daily Tweets with revelations of sexual abuse or assault and early life experience of sexual abuse and assault in Figure 1. In the first two days of the movement, 11-13% of Tweets revealed an experience of sexual abuse or assault and over half of them revealed that the experience of sexual abuse or assault happened in early life. As the movement carried forward in the first week, the proportion of revelations of sexual abuse or assault decreased to about 6%, but the relative proportion with early life experience of these traumatic events increased to over 80%.

From SVM classification, we estimated the number of novel English Language Tweets that would have had a revelation of sexual abuse or assault in that first week of the #MeToo movement. Based on these numbers and the 25-75% range of followers from users with such experiences, we calculated a lower bound on the number of Tweeter user who may have seen a first-person revelation range from 5,955,342 to 34,251,628.[5]

Discussion

In 2006, Tarana Burke, a civil rights activist, coined the phrase "Me Too" as way to raise awareness about the pervasiveness of sexual violence. The power of these two words, included in the dramatic wave of Tweets that occurred the week of October 14, 2017 from women attempting to support one another was uniquely confirmed in the context of social media, bolstering and extending this evolving movement of resistance and resilience among women.

The Twitter data suggests that #MeToo gained traction very quickly. To understand how this happened we needed data from the beginning of movements to examine the content of the

[5] We assume we have a lower bound for reach because we did not include Retweets and our algorithm has a somewhat higher false negative rate. However, we were unable to count unique followers and therefore our reach estimates may be elevated if the network of people who reveal events are highly overlapping. This is an area we intend to explore further.

conversation. Twitter data illustrate that beyond the initial media attention on high profile cases, involving women in the public eye (actresses etc.), the collective sharing of personal experiences among ordinary women have enormous potential to further causes like #MeToo. Beyond only mentioning 'MeToo', as was the call from the initial Tweet put forth by Alyssa Milano, women revealed intensely personal and traumatic experiences. Based on our models, 11% of novel Tweets in the first week of the movement revealed details about the writer experience of sexual abuse or assault and 5.8% were about early life experiences, making them even more alarming. In addition, given the connectivity possible on Twitter, we estimate that between 5 and 34 million Tweeter users may have been exposed to such revelations.

While we acknowledge that the publication of the expose about Harvey Weinstein in The New York Times and the New Yorker (Oct. 9th, 2017) was the initial spark and that having a highly connected person, such as Alyssa Milano, start the campaign enabled the spread of the movement rapidly, we highlight here that the everyday and graphic revelations were likely another key component in spreading the movement.

In this brief, we documented the arc type of the revelation on Twitter between Oct. 14-Oct 21, 2017 and the proportion of Tweets with such revelations as an example of how social media amplifies and enables social movements message. We focus on the revelations of traumatic events on public platforms like Twitter to highlight how the content of initial conversation can be particularly salient and lead to rapid spread of movements. These data suggest that the mass sharing of personal experiences of sexual abuse and assault, filled with deeply personal and painful narratives, has certainly contributed the strong and ongoing #MeToo movement we now observe.

Table 1: Words contained in Tweets that consistently predict a revelation of sexual abuse or assault.

MOST PREDICTIVE WORDS CATAGORIZED					
TIME/AGE	PERSON	ACTION	BODY PART	INTOXICATION	OTHER
Age	Boyfriend	Advantage	Arm	Drug	Daylight
College	Coworker	Chase	Boob	Drunk	Door
First Time	Date	Grab	Butt	Frat**	Interview
Fifteen	Father	Grope			
Grade	Man	Rape			
Hasnumber*	Rapist	Shout			
Kindergarten	Police				
Old	Stranger				
Year	Teacher				
Years Ago	Uncle				
Years old					

* Hasnumber is an overall indicator of whether there is a number in the text
** This term can indicate location, time, an implied level of intoxication, etc.

Table 2: Words contained in Tweets that consistently predict a revelation of an early experience of sexual abuse or assault.

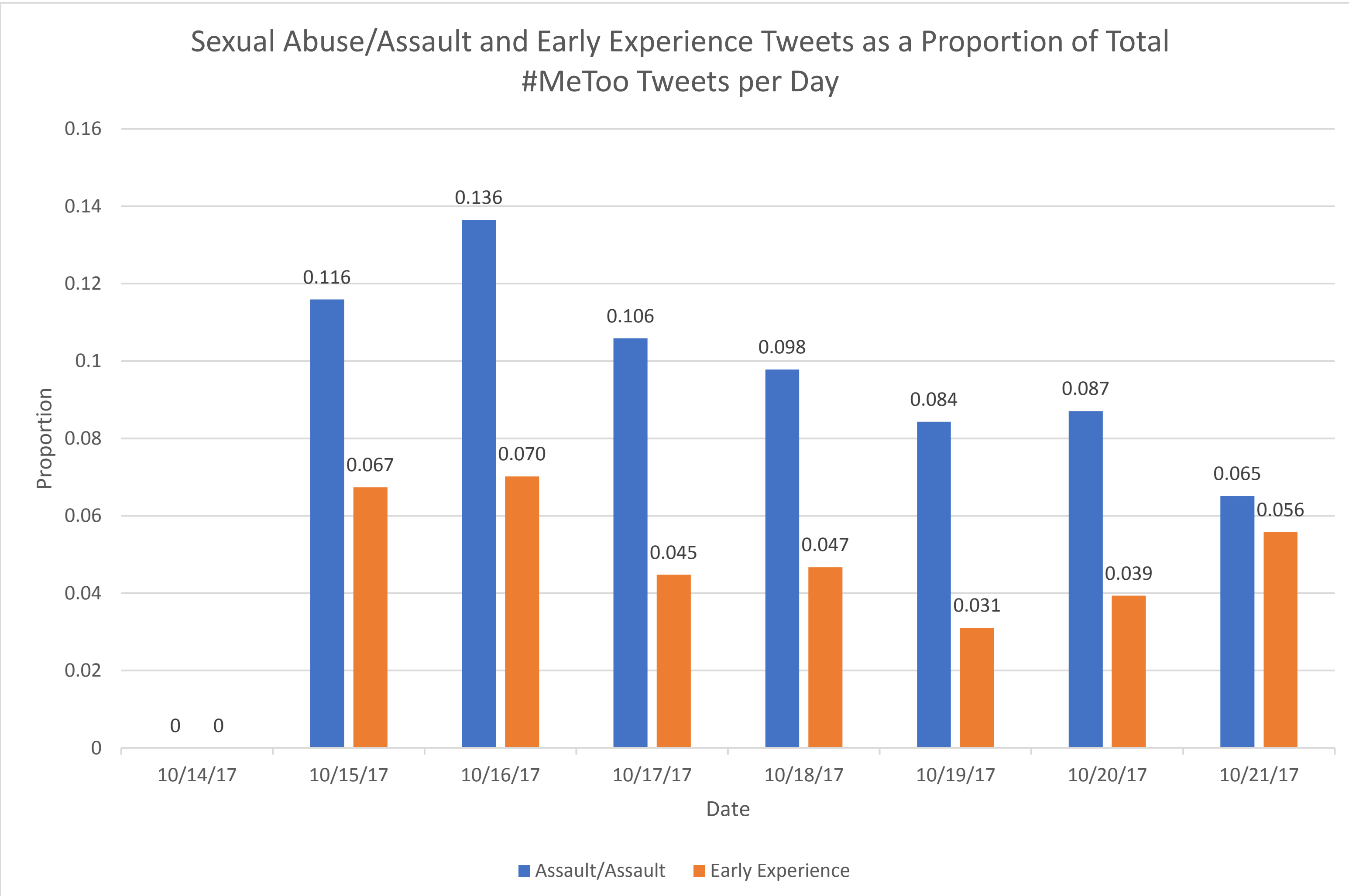
MOST PREDICTIVE WORDS CATAGORIZED					
TIME/AGE	PERSON	ACTION	BODY PART	INTOXICATION	Other
Age	Cop	Afraid	Arm	Drunk	Concert
College	Coworker	Rape	Butt		
First Time	Doctor	Rip	Pussy		
Freshman	Father				
Grade	Male				
Hasnumber*	Neighbor				
High School	Rapist				
Kindergarten	Step Father				
Old	Teacher				
School	Uncle				
Years Ago	Date				
Years Old					

* Hasnumber is an overall indicator of whether there is a number in the text

Table 3: Estimates of Number, Proportion and Reach of MeToo Tweets from Oct. 14-Oct. 21, 2017

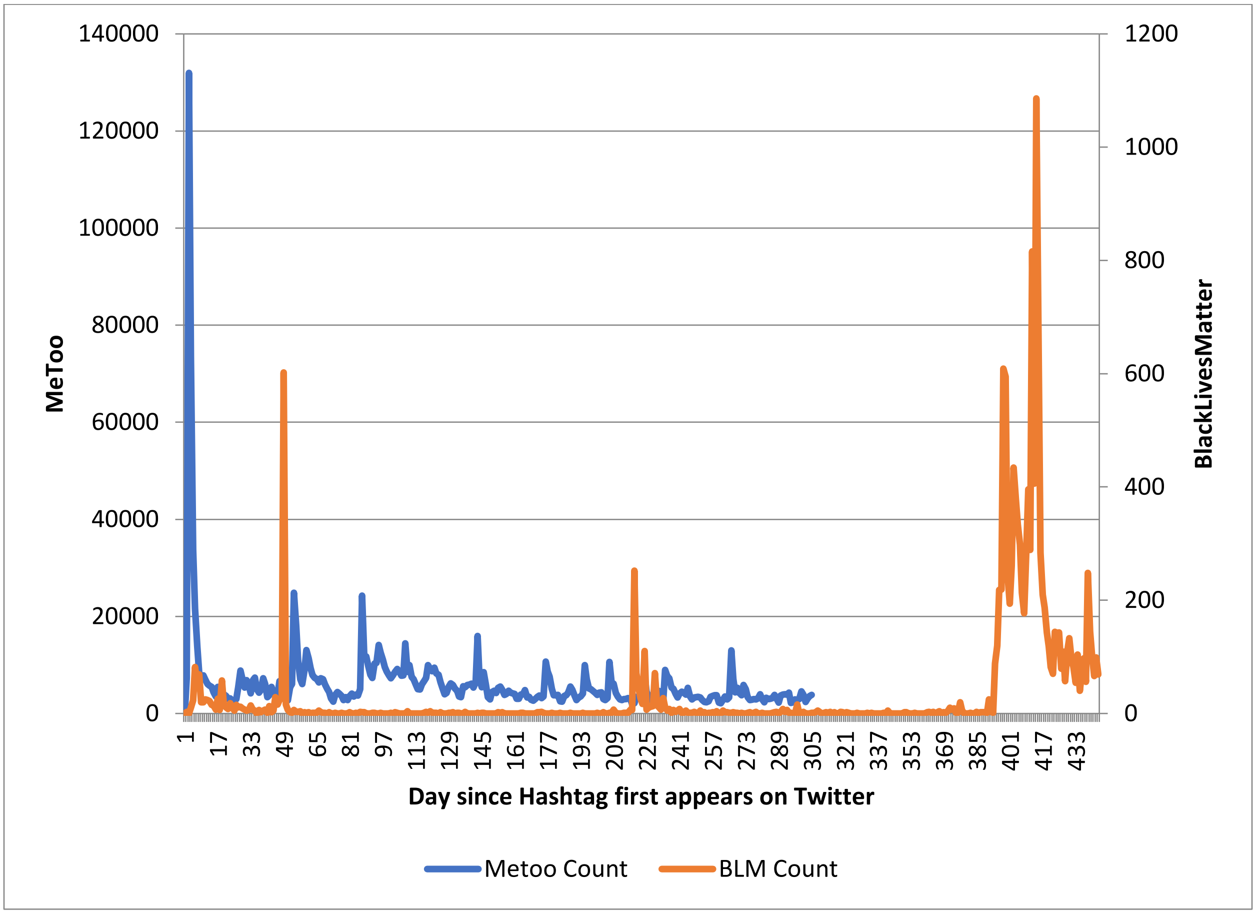
Tweet Type	Tweets Categorized by SVM	Percent of Novel English Language Geotagged Tweets	Est. of Novel English Language Tweets	Lower Bound Estimates of Potential Reach
Assault/Abuse	1287	11.4%	33,646	[5,955,342 to 34,251,628]
Early Experience	657	5.82%	17,176	[2,919,920 to 17,296,232]

Figure 1: Sexual Abuse/Assault and Early Experience Tweets as a Proportion of Total #MeToo Tweets per Day.



Note: Based on the analytical sample.

Appendices:
Appendix 1: Comparison of relative time patterns of Novel English Tweets including MeToo and BlackLivesMatter



Note: These are the daily novel English Tweets starting from one day prior to the creation of each hashtag on Twitter (July 13, 2013 for BlackLivesMatter and October 14, 2017 for MeToo). MeToo counts are on the left axis and BlackLivesMatter counts are on the right axis.

Appendix 2: All vs. novel Tweets—example table

MeToo Tweets Examples	Novel Me Too Tweets Examples
"RT @500daysofMary: I was 13, he was 19. He lied to me and said he was 17. I was a child, but I was told it was my fault. #MeToo https://t.c… "	"Every women in my family #MeToo has been raped. What are our national stats and how do we compare? #IHearYou"
"RT @womensmarch: To all the women sharing stories of sexual assault and sexual harassment, thank you for your bravery to speak up. You are not alone."	"#metoo A 'friend' took 'care' of me after a party in college because I drank too much. Care = Rape I was 18. I trusted him."[A, B]
"RT @MarisaKabas: my entire twitter & facebook feeds are full of women i know saying #metoo. Men, no matter what your history---just let this sink in."	"First time, 1995 when I was dead asleep and woke up to a man forcing himself on me. Too scared to fight, apparently NO wasn't enough. #meToo" [A]

Note: Tweets are presented in their original form as typed by the original poster.

[A] Example of sexual assault or abuse Tweet
[B] Example of early life experience of sexual assault or abuse Tweet

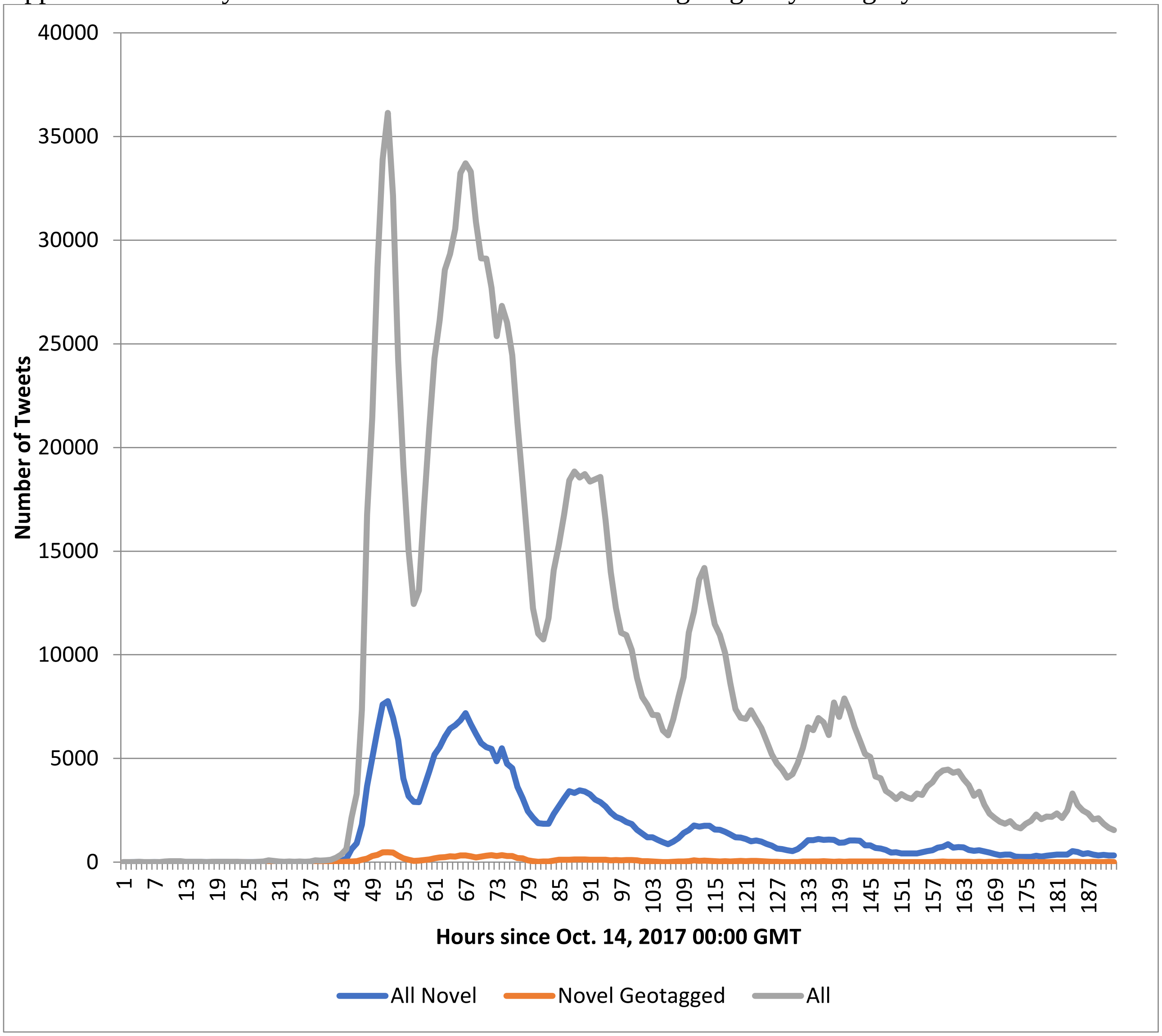
Appendix 3: Detailed Data Extraction Description

Several factors pertaining to obtaining Twitter Premium API data contribute to the discrepancy between the 12,337 novel English geotagged Tweets and the subsample sample of 11,935 novel English geotagged Tweets that we use in this study. The Twitter Premium API has a maximum limit of 500 Tweets per data request and each request costs money. Therefore, we had to optimize our requests.

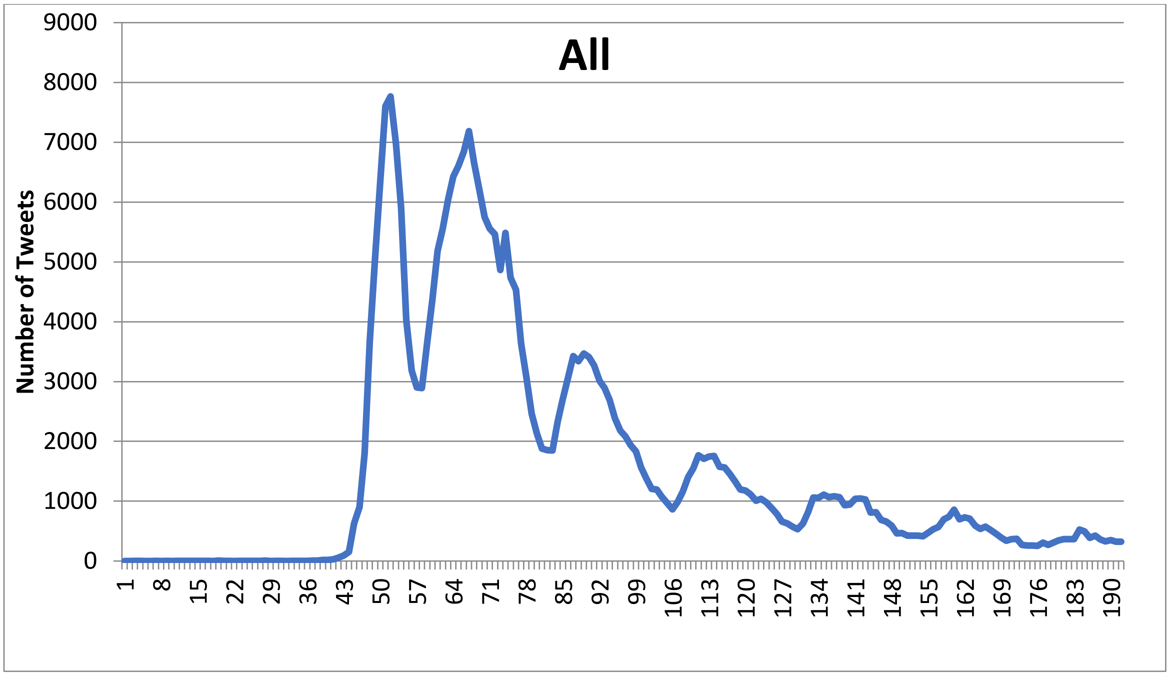
We decide to downloaded Tweets based on non-overlapping times in which they originated. This process led to the loss of some Tweets during periods with more than 500 Tweets. For example, if we requested Tweet between 8pm-9pm on Oct. 15 and there were over 500 Tweets, say 637, we recovered only the first 500. The remaining 137 Tweets would have to be recovered in a separate request. Therefore, we examined the data to try to ascertain the time of last tweet downloaded to then send a new request with non-overlap times to fill in missing Tweets. However, this process was inherently error prone. In some cases, we lost a few minutes and in others we overlapped in time. This led to both some duplicate Tweets and brief periods with missing data.

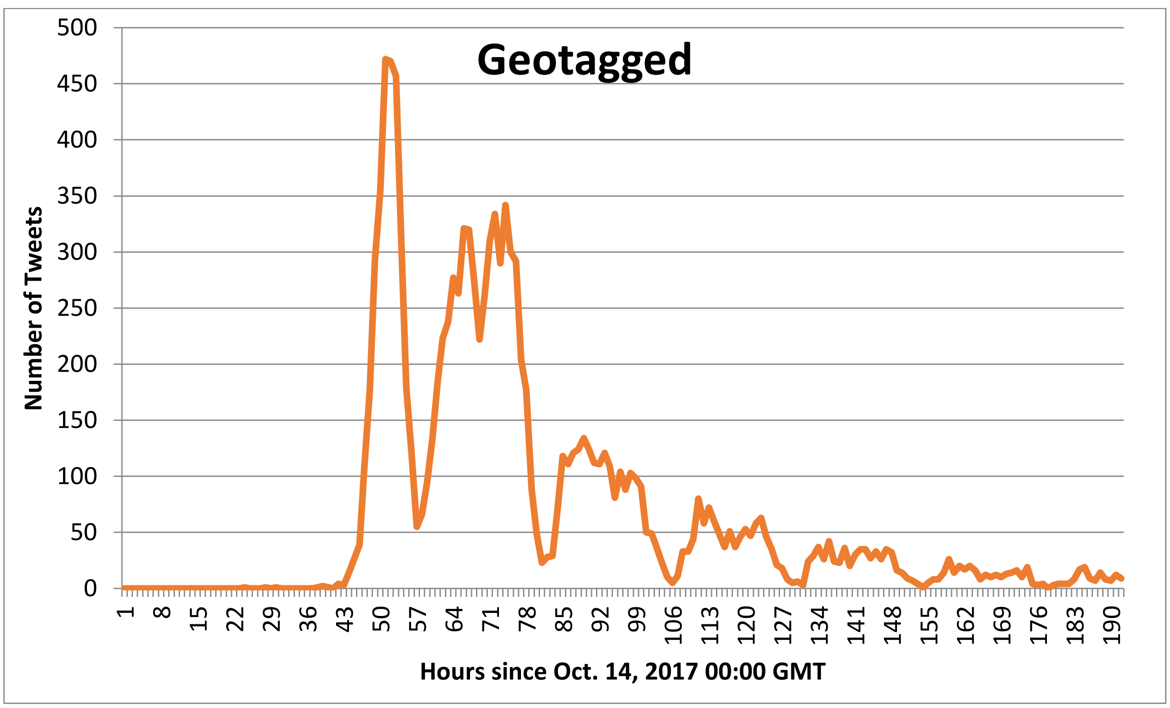
After data cleaning we had 11,935 non-duplicates Tweets but based on API there were 12,337 that should have met our search criteria.

Appendix 4: Hourly counts of MeToo Tweets from hashtag origin by Category



Appendix 5: Hourly count of MeToo Tweets detailed comparison of All vs. Geotagged





Appendix 6: Top 20 locations of Tweets in analytical sample

City	**Count**	**City**	**Count**
New York City, NY	714	Philadelphia, PA	84
Los Angeles, CA	602	Columbus, OH	77
Chicago, IL	274	Phoenix, AZ	66
Washington, DC	240	Denver, CO	63
Bay Area, CA	204	Dallas, TX	63
Austin, TX	125	Atlanta, GA	59
Seattle, WA	124	San Antonio, TX	59
Houston, TX	122	Nashville, TN	58
Portland, OR	119	Minneapolis, MN	52
Boston, MA	98	Omaha, NE	42
San Diego, CA	95	Kansas City, MO	41