5-2018

Explaining Animosity Towards the Roma: A Case Study of Twitter Communication in Italy during the Refugee Crisis

Mayuko Nakatsuka

The Graduate Center, City University of New York

EXPLAINING ANIMOSITY TOWARDS THE ROMA

A CASE STUDY OF TWITTER COMMUNICATION

IN ITALY DURING THE REFUGEE CRISIS

by

MAYUKO NAKATSUKA

A master's thesis submitted to the Graduate Faculty in Liberal Studies in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2018

Explaining Animosity towards the Roma

A Case Study of Twitter Communication in Italy during

the Refugee Crisis

by

Mayuko Nakatsuka

This manuscript has been read and accepted for the Graduate Faculty in Liberal Studies in satisfaction of the thesis requirement for the degree of Master of Arts.

______________________ ____

Date Till Weber

Thesis Advisor

______________________ ____

Date Elizabeth Macaulay-Lewis

Acting Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Explaining Animosity toward the Roma

A Case Study of Twitter Communication in Italy during the Refugee Crisis

by

Mayuko Nakatsuka

Advisor: Till Weber

Italy is known for hostile treatment of the Roma, one of the largest ethnic minority groups in Europe. This paper seeks to understand what is causing Italians to talk negatively about the Roma on Twitter. Statistical analysis is performed utilizing the data mined from Twitter along with other variables. The study finds that Roma population, foreign population, and number of refugees all have significant effects on the total number of tweets or the average negative sentiment of tweets. The results indicate that native Italians may group minority groups all together and regard them as “others”. Although the research design has some flaws in the data mining and sentiment analysis process, the study shows promise. I suggest that social scientists utilize social media data to analyze social or cultural phenomena.

TABLE OF CONTENTS

LIST OF FIGURES

\

LIST OF TABLES

Introduction

According to a recent study published by Pew Research Center (2016), 82% of Italians have negative opinions about the Roma, one of the largest ethnic minority groups in Europe. This is the highest percentage out of all the countries surveyed; Greece, Hungary, France, Spain, Poland, the United Kingdom, Sweden, Germany, and the Netherlands. The median was 48% (Pew Research Center 2016). Greece and France, the two countries known for open hostility toward the Roma, scored 67% and 61%, respectively, well below what Italy did (Pew Research Center 2016). The same study also found that Italians are most critical of Muslims, with 63% of them possessing negative opinions (Pew Research Center 2016). Although 63% seems like a high percentage at a first glance, it is considerably smaller than 82%, the percentage of Italians who have negative opinions about the Roma (Pew Research Center 2016). Why does such a large percentage of Italians possess negative views on the Roma? How are their views influenced by the arrival of refugees, the number of immigrants and the number of Roma living in their city? Do their views differ from one city to another? These research questions will be addressed on the basis of a quantitative analysis of communication on *Twitter*.

Social media has become part of many people's lives since Twitter, Facebook, and Instagram were introduced and gained popularity all over the world. Even those who do not use social media are exposed to social media through other forms of media, such as television and newspapers. As of January 1, 2018, Twitter has about 330 million active users (Aslam 2018). Around one third of them tweet every day, and a little over 80% of them use their mobile devices to send and read tweets (Aslam 2018). About 11% of the Italian population uses Twitter (Russo 2017). The platform is very easy to navigate and makes it possible for anyone with internet access to express their thoughts in a couple of short sentences and share them with a large audience. They write freely and interact with other users by retweeting[1] replying and favoriting[2]. Retweeting exposes tweets to more people and spreads their content beyond the user's followers.

Many users do not seem to shy away from expressing negativity or hateful feelings for social and racial minority groups on Twitter, where the users do not have to communicate directly with the subjects face-to-face. Although users have the option to make their accounts private, only about 10% of them are private, and the rest of them choose to keep their account public (Bosker 2012). Timelines or individual tweets of public accounts can be embedded in websites and blogs, which helps expose tweets to a wider audience who does not use Twitter. It is unclear what percentage of users have chosen to reveal their real name on Twitter. However, it is easy to spot accounts that belong to those who have opted to hide their identity and use a pseudonym. Some seem to create accounts just to attack and harass others.

Although Twitter prohibits the promotion of hate content, sensitive topics, and violence globally, while allowing its users to block accounts and filter words, racist and derogatory taunts against social and racial minorities are floating around (Twitter). Since one can direct a tweet at another user,[3] personal attack against social and racial minorities often takes place as well. The company does not want to limit freedom of expression, which makes it even easier for its users to show hostility towards social and racial minority groups like the Roma.

---

[1] Repost or forward someone else's tweet.
[2] Click the heart icon of a tweet and let one's followers and the author know that s/he likes the tweet.
[3] Users can direct a tweet at another user by adding @username.

I figured that rather than surveying how many people, or in my case, Italians, think negatively of the Roma, collecting tweets pertaining to the Roma and measuring how negative/positive each tweet was over the span of a couple of years would be useful to understand why Italians view the Roma negatively, and what is affecting their views. Also, the total number of tweets, either negative or positive, can indicate Italian people's interest in the Roma. Turning people's sentiments regarding the Roma on social media into numbers seemed like the right way to gauge how Italians think of the Roma. Rather than interviewing millions of Italians in person who may be afraid of revealing their true feelings to a complete stranger, mining tweets seemed like such an efficient way to gauge public opinion on the Roma in urban areas in Italy. I set out to find what was influencing the number of tweets about the Roma and the sentiments of those tweets in the thirty largest cities in Italy.

This paper is divided into three chapters. In Chapter 1, I present a brief sketch of the history of the Roma people and their current situation in Italy. Chapter 2 explains how tweets were mined and analyzed. Also, relevant literature is examined. Chapter 3 is divided into two sections. The first section presents a cross sectional analysis of the influence of the Roma population and immigrant population on Twitter activity. The second section presents a time series analysis of the influence of refugee arrivals on Twitter activity.

Chapter 1. The Roma

1.1 History of the Roma People

The Roma are said to be of Indian origin, and currently live mostly in Europe and the Americas (Hancock 2002). Researchers present conflicted views on the origin of the Roma, but the majority of them seems to have agreed that they have roots in India (Hancock and Kyuchukov 2005). Today they are often referred to as gypsies by the media and general public although many Roma people consider the term gypsy to be derogatory and prefer to be referred to as Romani or Roma instead (Hancock 2002). Some still embrace the meaning behind the term gypsy and proudly use it to refer to themselves (Matras 2015). The Roma arrived in Europe in the twelfth century and did not settle down in one country (Hancock and Kyuchukov 2005). Although a few countries gave the Roma the right to travel and stay freely at first, by the sixteenth century many countries started enslaving or expelling them (Hancock and Kyuchukov 2005). In pursuit of basic rights, a large number of Roma people moved to Eastern Europe where they were not subject to ethnic cleansing (Achim 1998). Later in the nineteenth century, the Roma started immigrating to the Americas from England and Eastern Europe (Achim 1998).

During World War II, the Nazis committed genocide of the Roma in Europe. Many Roma call this horrific event "*porajmo*"[4] (Hancock 1997). It is estimated that the death toll of the Roma in World War II ranges from 220,000 to 1,500,000 (Hancock and Kyuchukov 2005; United States Holocaust Memorial Museum). The estimate varies greatly depending on researchers. Since the governments of Poland and Germany did not recognize the genocide of the Roma until 1982 and 2011, respectively, the Roma porajmo survivers did

[4] The term was coined by Hancock, but it is not an official term and some Roma prefer not to use it to refer to the genocide.

not receive any war reparations from these governments (The Telegraph 2011). Many Roma rights organizations have been campaigning to raise awareness about *porajmo*. Unlike Jewish people who have managed to build their own country and succeeded in many industries all over the world, the Roma are still dispersed and the majority of them has not succeeded economically. The overall low socioeconomic status does not even come close to that of Jewish people. Having a low socioeconomic status, the Roma do not even come close to Jewish people in terms of status in society.

1.2 Current Situation of the Roma People

Although they are one of the largest minority groups in Europe, the Roma are still discriminated against in contemporary society. Many of them live in camps with the fear of being evicted. The survey conducted by the European Union Agency for Fundamental Rights shows that over 30% of the Roma in Italy live in illegal settlements, which is the highest percentage in all the countries surveyed [5] (European Union Agency for Fundamental Rights 2011). The European Union has acknowledged the problem and encouraged each member country to take a step in the right direction. Moreover, the European Commission published the Roma Integration Strategy (2011) to improve the housing, education, employment, and healthcare of the Roma. However, no significant progress has yet to be reported (The European Union 2017).

Currently, it is estimated that about 150,000 Roma live in Italy, about half of which are Italian citizens (Aragona 2015). Since Romania joined the European Union in 2007, the number of Roma in Italy has been gradually increasing even though the Italian

[5] Italy, Greece, Czech Republic, Slovakia, Romania, Portugal, Spain, Poland, Bulgaria, France, and Hungary participated in the survey.

government started deporting Romanian residents with criminal records, most of which are Roma (BBC News 2007). Although many people have the misconception that the Roma are nomad, only 2-3% of the Roma and Sinti[6] people in Italy are nomads (USC Shoah Foundation 2013). As shown in the 2011 Roma Survey results, the Roma experience discrimination in many areas of their life, and 97% of them are at risk for poverty[7] (The European Union Agency for Fundamental Rights 2011). It is not an exaggeration to say that the Roma are one of the most vulnerable minority groups in Italy.

[6] The Roma of Central Europe.

[7] Their monthly income is below the national at-risk-of-poverty threshold.

Chapter 2

2.1 Methodology

2.1.1 Data Mining Method

The major challenges for social scientists who work with data obtained from social media are collecting and managing the data, turning the text into numbers of some sort, and analyzing the numbers (Carrigan 2014; Woodfield 2014). Textual analysis has long been around in social sciences and humanities as a research tool to analyze traditional media, and it can be both qualitative and quantitative. However, most social science research involving media and texts tends to be qualitative. This can be attributed to the lack of training in data mining and management. Like many social scientists, I have limited knowledge of programming. In preparation for this project, I have acquired the basic skills necessary for data mining and management. With basic coding abilities and some experience in data management, one can achieve new research techniques that can boost his or her research significantly.

I wrote codes on Python that connect to Twitter's application programming interface (API) and mined live tweets that include keywords related the Roma in English and Italian: Zingari, Zingaro, Camminanti, Camminante, Nomade, Nomado, Rom, Roms, Roma, Romas, Romani, Gypsy, Gypsies, Gitano, Gitani, Sinti, and Sintis. The program returns each tweet with username, time, date, language, geotag (location information) if any, and other details in JSON format. The access key and the password to the Twitter API were obtained through my personal Twitter accounts. Since the Twitter API often disconnects the program if it keeps running for more than a day or two, the program was made to restart at midnight EST every day. Technically, my dataset is missing tweets from

23:59:59 to 00:00:00 EST each day, which means I missed less than 1/86400 of the total tweets. I consider this problem a minor one because not so many people in Italy tweet right around 6am their time. The program was run on a server from January 1, 2015 to December 31, 2017.

I downloaded data from my server periodically and sorted tweets. After tweets were collected, I selected those with a geotag from one of the thirty largest Italian cities: Rome, Milan, Naples, Turin, Palermo, Genoa, Bologna, Florence, Bari, Catania, Venice, Verona, Messina, Padua, Trieste, Taranto, Brescia, Prato, Reggio Calabria, Modena, Parma, Perugia, Reggio Emilia, Livorno, Ravenna, Foggia, Cagliari, Rimini, Salerno, and Ferrara. These tweets amount for about 20% of all the tweets collected. Then, many tweets in Italian that had keywords related to sightseeing or sports were also deleted from the list. Rome is spelled Roma in Italian, which caused my program to mine tweets that had nothing to do with the Roma. "I live in Roma," "I'm in Roma," "I'm going to Roma," etc., were also deleted.[8] Although some Italians or non-Italian residents seem to write their tweets in English, I made the decision to exclude tweets in English even if they were accompanied by a geotag of one of the aforementioned Italian cities so that sentiment analysis would be uniform. If I were to measure the sentiment of those tweets written in English, I would have to use another "bag of words," which might have a different scale for measuring sentiment. This means that tweets written by tourists and non-Italian residents who do not speak Italian are excluded from my dataset, which may not reflect the view of local non-Italian residents.

[8] I had help from two native Italian speakers. Since it was impossible to go through each tweet, this method is not perfect.

2.1.2 Sentiment Analysis

Computer engineers Balasubramanyan, O'Connor, Routledge, and Smith examine the relationship between public opinion and sentiment measured from text on Twitter (2010). Their analysis found that public opinion correlates with sentiment word frequencies in tweets. In several cases, the correlations are as high as 80%, and capture important large-scale trends (Balasubramanyan et al. 2010). The results of the research highlight the potential of text streams as a substitute and supplement for traditional opinion polling that is time and money consuming.

To analyze sentiment of each tweet, the "bag of words" approach was used. I downloaded a set of words (lexicon) with quantified sentiment values in Italian from GitHub, which was made available to the public for free by the Italian computer scientist Andrea Cirillo. I am not fluent in Italian, so I could not check the accuracy of his lexicon myself. In order to test Cirillo's lexicon, I had two native Italian speakers take a look at it and test its accuracy. Since both of them approved of the way the lexicon worked, I made the decision to continue using it. Sentiment of each tweet was measured on a scale from zero to one. Zero means the most positive, and one means the most negative.

I am well aware that connecting to a natural language processing toolkit (NLPK) API would have been the best option to perform more precise sentiment analysis, but my limited funds did not allow me to use an NLPK API to process the very large dataset. Still, the lexicon used for this study was good enough to analyze tweets because the dataset is relatively simple due to the fact that Twitter used to have a character limit of 140 for each tweet. Users tend to write their tweets using simple words and sentences. In November 2017, the company doubled the character limit to 280 characters (Larson 2017). I have taken a look at my dataset from November to December 2017 and found that more than 90% of tweets still fit in the old character limit of 140. The simplicity of tweets made it possible to rely on the "bad of words" approach.

2.2 Current Issues

2.2.1 Difficulties in Data Collection

As pointed out by leading Roma scholar Yaron Matras repeatedly in *Romani Studies Journal* (2009, 2010, and 2012), studies and research on the Roma tend to be qualitative. For now, there is a gap that needs to be filled by future studies. In their report on the condition of the Roma, the European Commission expressed regrets for not being able to collect accurate and reliable data since not all the Roma are registered in the country of their residence, and some of them refuse to identify themselves as Roma (The European Commission 2012). Moreover, even Roma who do not wish to be nomadic anymore are often forced to keep moving around in search of economic opportunities (The European Union Agency for Fundamental Rights 2011), increasing the difficulty of obtaining reliable data. This is one of the biggest challenges researchers face. In articles and data, we often see numbers that are only rough estimates. Also, it is not uncommon to see drastically different numbers from different sources. Researchers need to ask themselves what is it that is causing those numbers to be different.

Although it is unknown how many Roma from Romania are living in Italy, Romania has sent the largest number of immigrants to Italy between 2007[9] to 2017 (The Italian National Institute of Statistics 2018). Other than Romania, Ukraine, Serbia, Kosovo, Montenegro, Macedonia, Bulgaria, and Russia are said to have been sending their Roma

[9] Romania joined the EU in 2007. The Romanian population in Italy doubled in 2007.

citizens to Italy (Roma Europe 2012). According to European Roma Rights Centre, there are not enough resources and tools to examine and measure the number and the living conditions of them. Despite these difficulties coming from the unreliability of the data, I still expect to obtain meaningful results.

For this study, the tweets collected by the aforementioned program on Python from January 1st, 2015 to December 31st, 2017 were analyzed. The Roma population in each city was provided by the European Roma Rights Centre. The population and foreign residents of each city were taken from the Italian National Institute of Statistics website. The Italian National Institute of Statistics publishes reports on demographics every year. However, I was not able to find any reliable statistic on the Roma population in each city. For this project, I utilized the data provided to me by the European Roma Rights Centre, which is an estimate from 2012.

2.2.2 Gap in Literature

Analyzing negative sentiment of the Roma on Twitter is worthwhile because few social scientists have addressed the subject in this way. I chose the journal articles and books abstracted herein for their focus on issues around racism towards the Roma in Italy, and methodology for analyzing data from social media, notably Twitter. These articles and books are examined in order to help conduct primary research which studies the relationship between the number of tweets on the Roma or the sentiments of those tweets in the major cities in Italy and other independent variables such as the population of the Roma or the population of immigrants in each city.

There is no lack of literature illustrating the relationship between Italians and the Roma in general. In the process of making a book to show the current state of the Roma in

Europe, researcher Lucy Orta explores the poor living conditions of Roma camp residents in Italy and connects them to the local government's effort to keep the Roma away from the other desirable residential areas (2010). Accompanied by photographs, the poor living conditions of the camp residents are illustrated. Most of the camp residents hail from Eastern European countries, namely Romania (Orta 2010). However, there are some who hold Italian citizenship (Orta 2010). The interviews of non-Roma in the Florence metropolitan area reveal strong feelings of hatred toward the Roma even when they cannot come up with a particular reason or have never interacted with a Roma (Orta 2010). Orta questions why the Roma have to deal with the negative stereotype even more than other minority groups, such as Africans and Muslims. At the end, she emphasizes that the situation is not unique to Florence. In many other major cities and their outskirts in Italy, the Roma receive horrible treatment from local governments and are discriminated against (Orta 2010). This book is very significant in the way that it demonstrates how even some Italians who have never personally been inconvenienced by the existence of the Roma hold strong negative views of them.

In "Welcome 'in'. Romani migrants and Left-wing Tuscany (1987-2007)" (2011), scholar Giovanni Picker discusses the rationale behind the construction of nomad camps in Tuscany and examines the laws addressing the Roma created by Left-wing politicians. Picker argues that in the name of "cultural protection" of the Roma, the politicians pushed for segregation of the Roma from Italians and other immigrant groups (2011). The article emphasizes the significance of the role the Tuscany Left-wing politicians played in establishing a clear border between "us" and "them." Since the construction of the camps in the 1980s, Picker claims, discrimination against the Roma has been justified on the basis of their peculiar nature and needs that the local government assumed without consulting the Roma community (2011). Picker implies that other cities in Italy followed in Tuscany's footsteps and created laws that keep the Roma away from the areas that native Italians consider desirable. An article written by David Leask (2016) confirms that local governments all over Italy have continued to treat the Roma the same way, and not much has been improved.

There is some literature that suggests the usefulness of tweets in social science research. Computer Scientists Stefan Stieglitz and Linh Dang-Xuan examine emotions in tweets collected, and conclude that Twitter may be capable of reflecting collective emotive trends and thus has predictive power with regard to political events (2013). Stieglitz and Dang-Xuan suggest that social scientists, especially social psychologists and political scientists, try data mining and sentiment analysis as one of their research tools even though what artificial intelligence can do for us in terms of sentiment analysis is limited (2013). If social scientists are willing to learn or have help from those who are proficient in programming languages, data mining and analysis can indeed be a powerful tool to examine new social phenomena.

There has been collaboration between social scientists and computer scientists in efforts to analyze cyber hate. Pete Burnap and Matthew L. Williams analyze the data collected from Twitter following the 2013 murder of drummer Lee Rigby by two Muslim men of Nigerian descent in London, and measured how fast cyber hate of Muslims, Nigerians, and Africans spread (2015). Although they are not certain about the generalizability of the classifier or the statistical model used, researchers can still take hints from the model and collect data on social media following an event prompting a hateful homophobic response. A detailed comparison of sentiments and how hate spreads between a few events may be useful.

By taking hints from research conducted by computer scientists, I begin my empirical work to fill the gap in the literature on the root of hostility towards the Roma in Italy. It is hypothesized that one or more of the independent variables such as the number of immigrants and the number of Roma residents affect the number of tweets and the negative sentiment of tweets about the Roma in Italy.

First of all, here should be more chances for native-Italians to interact with and see the Roma when there is presence of a large Roma community nearby. They may see the Roma on the streets and have a conversation. They may hear or read about the Roma in local news or hear their neighbors talk about the Roma often. All these things should lead people to talk more about the Roma on social media.

Since Italians seem to have a strong sense of who they are and what "others" are, it is possible that minorities are grouped together as "others" by Italians. Thus, the independent variables not related to the Roma (the number of immigrants and refugee arrivals) could also be affecting the number of tweets and the negative sentiment of tweets about the Roma in addition to the independent variable pertaining to the Roma (the number of the Roma). When Italians feel threatened by immigrants in general or a certain ethnic minority group, they may feel that the Roma are part of that group that present a threat and talk more negatively about them.

My hypotheses are tested in the next chapter by examining the contribution of each independent variable (the number of the Roma, refugee arrivals, the total population, and the immigrant population) to the total number of tweets and the average sentiment of tweets about the Roma.

Chapter 3 Analysis and Results

3.1 Influence of the Roma and Foreign Populations on Twitter Activity

In this section, I study how the total population, the Roma population, and the foreign population affect Twitter activity regarding the Roma in each of the thirty largest Italian cities. More specifically, I focus on the total number of tweets and average sentiment of tweets about the Roma. The goal of this study is to measure the contribution of each independent variable to the models, and to provide statistical information regarding the confidence in those results.

3.1.1 Linear Model, Regression, and Hypothesis Rejection

The different variables involved in this study are Roma population, foreign population, total population, number of tweets, and average sentiment of tweets in the thirty largest cities in Italy. In order to quantify better the relevance of Roma population and foreign population within a city, the relative populations to the total will be considered. I am going to call P total population, R Roma population, and F foreign population. Percentage of Roma population and percentage of foreign populations are written as $r = 100R/P$ and $f = 100F/P$, respectively. For reasons of clarity, I am going to use the variable $p = P/1000$ (population per thousand). The independent are (r, f, p). The dependent variables are N, number of tweets within a city, and S, average sentiment of tweets within a city.

It is the goal of this study to establish the existence of an influence of the independent variables (r, f and p) on the dependent variables (N and S). This influence could be modelled in a great variety of ways. For the sake of simplicity, and in absence of any reason not to do so, a linear regression model will be used to perform cross-sectional analysis of the data.

In such a model, number of tweets and average sentiment can be written as

$$N(r, f, p) = n_r r + n_f f + n_p p + n_0$$

and

$$S(r, f, p) = s_r r + s_f f + s_p p + s_0.$$

In order to draw conclusions about this model, a multiple regression analysis using Stata is performed. This provides, through an ordinary least squares fit, the values of the parameters (n_r, n_f, n_p, n_0) and (s_r, s_f, s_p, s_0) that best describe the data. I call (n_r, n_f, n_p) and (s_r, s_f, s_p) the coefficients, while n_0 and s_0 are called offsets. Stata also provides multiple outputs that give an idea of the coherence between model and data, and some statistical information regarding the parameters.

In order to establish the existence of a multiple correlation between the independent variables and the dependent variables, within the linear model used, one has first to determine whether the data is properly described with the proposed linear model. To do so, one can use the coefficient of determination (or R-squared, or R^2). This number, ranging from zero to one, takes values close to one when there is a high correlation, and close to zero when there is almost no correlation. A high value of R-squared is a sign of the linear model being a good description for the data being analyzed.

After this, the obtained coefficients should be studied. More specifically, one has to study the departure of the coefficients (n_r, n_f, n_p) and (s_r, s_f, s_p) from zero. If any of those coefficients are zero, the dependent variable (N or S) would not have dependence on the respective variable. As an example, if $n_r = 0$, it would mean that the number of tweets is the same if Roma population is 20% or 80%. If n_r is positive, one can state that the number of tweets increases with the percentage of the Roma population. Conversely, if n_r is negative, one can state that number of tweets decreases with percentage of the Roma population. This applies to all the coefficients. The offsets represent the value of the dependent variable when the independent variables are zero. The expected values for n_0 and s_0 are, then, zero.

This part of the statistical analysis consists, then, in rejecting the possibility of the values of the coefficients to be zero. These possibilities (the coefficients being zero) are considered as the null hypothesis. The possibilities that one (or more) of the coefficients are not zero, are considered as alternative hypotheses. The output of the statistical analysis is whether the data allows or not the rejection of the null hypothesis, when confronted with the alternative hypotheses. This will be addressed in two independent ways:

a) Considering as the null hypothesis that all the coefficients are zero. This would mean that there is no dependence at all of the dependent variables on the independent ones, since $N(r,f,p) = n_0$ and $S(r,f,p) = s_0$, being both constant.
b) Considering as the null hypothesis that one of the coefficients is zero. This would mean that there is no dependence of the dependent variables in one of the independent variables, as explained above. This would be hypothesized for each of the three coefficients, related to the percentage of the Roma population, the percentage of foreign population, and the total population.

In total, four hypotheses would be considered for each model (N or S). The null hypotheses would be confronted with the alternative hypotheses, consisting in considering the full model with the parameters obtained from the least squares estimation. For the first null hypothesis (all coefficients are zero), a F-test is performed by the aforementioned statistical software. Another measure of the fitness of the complete model to the data is described by the R-squared, since a low correlation between the dependent and all the independent variables (low R-squared) would mean that the model is just a constant. For the three hypotheses regarding the individual parameters, a t-test is performed by the software. For each of the tests, we obtain a p-value, giving a significance of the rejection of the null hypothesis by the data. In order to reject the null hypotheses, one has to define beforehand (before looking at the data) a threshold used to reject the null hypothesis. This threshold is commonly called size of the test, or significance level. Specifically, if this threshold is α, the null hypothesis is rejected if $p < \alpha$. Usual values for α in social sciences, 0.1, 0.05 and 0.01, are going to be used here.

3.1.2 Number of Tweets

Here I discuss the results of the liner regression for the number of tweets model. The total number of tweets should indicate how much residents and visitors in each city are interested in the Roma and eager to talk about them on Twitter, either negatively or positively. It is important to note that none of the thirty cities surveyed had overall positive sentiment. A multiple linear regression shown in Table 2 and the statistical tests (the F-test and the t-test) reveal that percentage of Roma population, percentage of foreign population, and total population all are non-zero with high levels of statistical significance. The regression coefficients are positive for all of the aforementioned independent variables.

The null hypotheses (lack of correlation with one or more independent variables) are rejected at all significance levels.

Table 1

Total Number of Tweets Explained

% of Roma Population	200663.26**
	(18994.68)
% of Foreign Population	20354.73**
	(3649.72)
Population/1000	308.96**
	(24.25)
Constant	-42355.85
	(31067.03)
Observations	90
R-squared	0.84

+ $p<0.10$, * $p<0.05$, ** $p<0.01$

(OLS regression coefficient with standard errors in parentheses.)

An increase in any of those will likely cause the dependent variable, total number of tweets, to increase as well. For example, when population size increases by a thousand, we can expect to see around 309 more (mostly negative) tweets about the Roma. It is reasonable to expect total population to influence total number of tweets. Also, the significant effect that percentage of Roma has on total number of tweets comes as no surprise. The most important finding here is the effect that percentage of foreign population has on the total number of tweets even though it is not as large as that of percentage of Roma.

The R-squared is 0.84, which means that 84% of variance in the dependent variable (total number of tweets) can be explained by the independent variables (percentage of Roma population, percentage of foreign population, and total population/1000). The F-test for the whole model gives a p-value smaller than the three thresholds, so the null hypothesis is rejected.

The offset is $n_0 = -42355.85$, but this should not be understood literally. First of all, the total number of tweets will never be negative even though it can be zero. Moreover, although percentage of Roma population and percentage of foreign population can be zero, there is never a city with zero population. Thus, the value one would expect for this offset should be zero (or at least bigger than zero). The statistical test reveals, since the p-value is high, that it actually is compatible with the offset being zero.

3.1.3 Sentiment of Tweets

The average sentiment of tweets should indicate how negatively people talk about the Roma on average. Similar to the result described above for the total number of tweets, a multiple linear regression and the statistical tests reveal that all the independent variables are positive and non-zero with high levels of statistical significance. The null hypotheses (lack of correlation with one or more independent variables) are rejected at all significance levels except for s_f which is not rejected at the 0.01 level.

Table 2

Average Sentiment of Tweets Explained

% of Roma Population	0.159**
	(0.016)
% of Foreign Population	0.00724*
	(0.00287)
Population/1000	0.000118**
	(0.000021)
Constant	0.533**
	(0.027)
Observations	90
R-squared	0.62

+ p<0.10, * p<0.05, ** p<0.01

(OLS regression coefficient with standard errors in parentheses.)

Similar to the analysis of total number of tweets, an increase in any of those will likely cause the dependent variable, total number of tweets, to increase as well. For example, if percentage of Roma population increases by 1, we should see average negative sentiment increase by 0.159.

R-squared is 0.62, which means that 62% of variance in the dependent variable (total number of tweets) can be explained by the independent variables (percentage of Roma population, percentage of foreign population, and total population/1000). The F-test for the whole model gives a p-value smaller than the three thresholds, so the null hypothesis is rejected.

The constant is 0.533, and it is an indication of the expected sentiment that a small town without Roma population and foreign population would reach. It is interesting to observe that the constant is almost at neutral sentiment.

Table 3

Correlation between the Percentage of the Roma Population and the Percentage of Foreign Population

	% of Roma Population	% of Foreign Population
% of Roma Population	1.0000	
% of Foreign Population	-0.0705	1.0000

The percentage of the Roma population and the percentage of the foreign population have a low negative correlation and are thus virtually independent of each other. There are not always Roma communities or camps in cities with a large immigrant population. As shown in table 1, the foreign population has a non-zero value with a high statistical significance. Although the number of the Roma has a greater effect on the total number of tweets, we cannot overlook the effect of the foreign population on the total number of tweets. The more immigrants there are in a city, the more Italian Twitter users in that city talk negatively/positively about the Roma even if there is very little or no population around them.

Table 4

Mean Sentiment of Tweets by City

(0%=Positive, 50%=Neutral, 100%=Negative)

City	Mean Sentiment 2015	Mean Sentiment 2016	Mean Sentiment 2017
Rome	96%	94%	90%
Milan	93%	92%	94%
Naples	81%	77%	72%
Turin	81%	80%	81%
Palermo	73%	70%	70%
Genoa	64%	59%	58%
Bologna	65%	64%	63%
Florence	96%	95%	98%
Bari	60%	60%	65%
Catania	70%	68%	69%
Venice	68%	65%	65%
Verona	67%	66%	66%
Messina	67%	65%	67%
Padua	61%	60%	60%
Trieste	55%	55%	60%
Taranto	60%	60%	61%
Brescia	94%	91%	90%

Prato	62%	60%	65%
Reggio Calabria	69%	69%	69%
Modena	69%	68%	68%
Parma	68%	68%	68%
Perugia	59%	57%	57%
Reggio Emilia	74%	69%	69%
Livorno	80%	79%	82%
Ravenna	70%	68%	68%
Foggia	55%	54%	50%
Cagliari	68%	65%	65%
Rimini	58%	58%	58%
Salerno	66%	65%	64%
Ferrara	58%	57%	61%

All the thirty cities have neutral to strongly negative sentiment. As expected, cities known for unfair treatment of the Roma and the large presence of illegal settlements and camps such as Florence, Rome, and Milan show strongly negative sentiment (Aragona 2015).

3.2 Influence of Refugee Arrivals on Twitter Activity

Now we turn to a second analysis of the Twitter data. In the cross-sectional analysis performed above, it is difficult to tell whether the negative opinions on the Roma come from grouping of the Roma and other immigrants as the same entity of "others". Therefore, not we conduct a time series analysis. The refugee crisis is an excellent opportunity to test this because it featured fast and exogenous inflow of foreigners.

Along with Greece and Spain, Italy has been receiving refugees from Africa and the Middle East. Refugees risk their lives to cross the Mediterranean Sea. The United Nations High Commissions for Refugees (UNHCR) registers refugees in the three aforementioned countries and publishes statistics on their website on a regular basis. A total of 454,647 refugees arrived in Italy from January 2015 to December 2017 (UNHCR 2018). The numbers provided by UNHCR does not include economic immigrants from non-EU countries and refugees arriving by air. Since native Italians seem to group non-Italians as "others," it is worth examining the relationship between the number of refugees arriving by sea.

As can be seen in table 5 below, the dataset from 2015 in the middle of the Syrian refugee crisis shows stronger negative sentiments towards the Roma. Italy had been constantly receiving refugees from Africa, and numbers of Syrians joined them starting in 2014. Although Italy did not take as many Syrian refugees as Greece did, the country was hit with a new heavy burden and responsibilities (UNHCR 2018). In all the thirty cities, number of total tweets, mean (negative) sentiments, and number of negative tweets are greater in 2015 than they are in 2016. Number of total tweets decreased by about 13% in 2016. Even though the mean sentiment went up again in 2017, number of tweets kept decreasing. This comparison result indicates that it is possible that Italian Twitter users' views on the Roma were affected by the Syrian refugee crisis. Since the first country refugees arrive at has to process their asylum applications, along with Greece, Italy had to take a heavy burden. Yet, there is no record that shows an increase of Roma immigration during the refugee crisis. It is possible that a few large cities with a relatively small Roma population lost some interest after the peak of the Syrian refugee crisis. In order to test the hypothesis that Italian Twitter users' level of interest in the Roma and how they perceive the Roma is affected by current political or social events that are not directly related to them, constant monitoring of Twitter activities and how they change as certain types of events occur is necessary.

Table 5

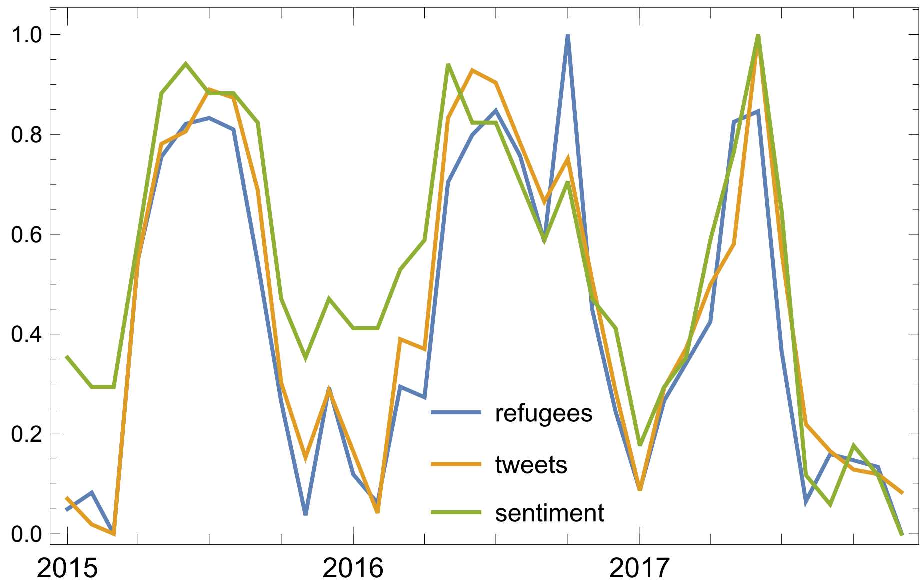
Overview

	2015	2016	2017
Number of Total Tweets	11536609	10014552	9625179
Mean (Negative) Sentiment	70.07%	68.80%	69.92%

As a first indication that there may be a relation between refugee arrivals and Roma related twitter activities, one can see in figure 1, the curves for the evolution of refugee arrivals, number of tweets, and sentiment of tweets exhibit similar features during the three-year period. Notice that the three quantities have been linearly scaled to range from zero to one between their minima and maxima.

Figure 1

Evolution of Refugee Arrivals, Number of Tweets, and Sentiment



Evolution of refugee arrivals, number of tweets, and average (negative) sentiment.

3.2.1 Linear Model, Regression, and Hypothesis Rejection

When there is time series data to be analyzed, the interest is in studying the structure of the data as time passes. In the present case, the structures to be studied are the relation between the number of tweets and the number of newly arrived refugees, as well as between the sentiment of tweets and the number of newly arrived refugees. Moreover, the yearly structure of the dependent variables can be understood by conducting this analysis. This reflects that refugees prefer warmer months to cross the Mediterranean sea. To do so, this analysis mixes numerical variables (such as sentiment S or number N of tweets, or the number of refugees g) and categorical variables (such as the month of the year, m). The linear regression model has to be slightly altered to treat categorical variables.

In general, if the categorical variable can take n different values, one needs to introduce $n-1$ new variables in the linear model, with their correspondent $n-1$ parameters. One of the possible values doesn't have an associated variable; this is called the base value. In my study, the categorical variable is the month of the year. Although different choices of base values produce different numbers in the linear model, the predictions for the dependent variable do not change. For simplicity, here January is selected as the base value. Then, I introduce, eleven variables, corresponding to the eleven months from February to December. These variables are called X_k. I introduce, as well, eleven parameters, called b_k. The label k can go from two to twelve and represents the number of the month. The variables X_k are kind of intermediate variables between the independent variable m (month of the year) and the dependent variable N (number of tweets). Indeed, the variables X_k are dependent on the variable m; thus, I write them as $X_k(m)$. If one considers the month February, with $m = 2$, the eleven intermediate variables are specified by $X_2(2) = 1$ and $X_k(2) = 0$ for $k = 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$. The linear model can then be written as

$$N(g, m) = n_g g + \sum_{k=2}^{12} b_k X_k(m) + n_0.$$

To clarify this, in the February example, the number of tweets would be given by

$$N(g, 2) = n_g g + \sum_{k=2}^{12} b_k X_k(2) + n_0 = n_g g + b_2 + n_0,$$

since all the $X_k(2)$ for $k \neq 2$ are zero.

Once the model is constructed, the multiple regression by least squares and the statistical analyses are the same as in the previous cross-sectional analysis part. The way to extract information about the correlation between number or sentiment of tweets and number of refugees is the same as before as well, since the number of refugees is an analogous variable to number of Roma people, foreign people or total population. Regarding the information in the parameters associated which each month, the procedure is a bit different. For a particular month m, the number of tweets is

$$N(g,m) = n_g g + b_m + n_0.$$

For the base month, January ($m = 1$), the number of tweets is

$$N(g,1) = n_g g + c.$$

Then,

$$b_m = N(g,m) - N(g,1).$$

This means that the values of the coefficients b_m measure the difference between the month m and the base month, January. Studying the evolution of the coefficients b_m throughout the year (for all values of m), one can study the yearly modulation of the number of tweets that is not caused by the arrival of refugees.

In a situation where most of the evolution of the number of tweets is explained by the arrival of refugees, the value of n_g is expected to be non-zero, and the values of the coefficients b_m are expected to close to zero, since they measure the monthly variation with respect to January. The null hypothesis we want to reject is as follows: there is no correlation between the number of tweets and refugee arrivals, and all the changes in the number of tweets can be explained as a seasonal effect. Numerically, the null hypothesis is that $n_g = 0$ and $b_m \neq 0$. Rejection of the null hypothesis will be achieved if the statistical tests give a p-value for n_g smaller than the thresholds, and the p-values for the

b_m are high. The same reasoning applies to the sentiment of tweets and s_g, where the linear model is

$$S(g, m) = s_g g + b_m + s_0.$$

3.2.2 Number of Tweets

Here I present the results and analysis of the studies of the correlation between refugee arrivals and number of tweets about the Roma in the thirty largest cities in Italy. In table 6, I show the values of the parameters of the linear model considered above.

Table 6

Number of Tweets in Time Series Explained

Parameter	Value
Refugees	53.50**
	(4.65)
January	0.00
	(.)
February	-52499.92
	(109572.93)
March	69160.05
	(110402.73)
April	153581.04
	(116007.66)
May	111969.25
	(134898.92)
June	324558.10*
	(139160.18)
July	308531.32*
	(129689.91)
August	230594.20+
	(121808.38)
September	189901.78
	(116549.78)
October	-48721.34
	(118286.62)
November	86576.06
	(110327.75)
December	55402.93
	(109963.02)

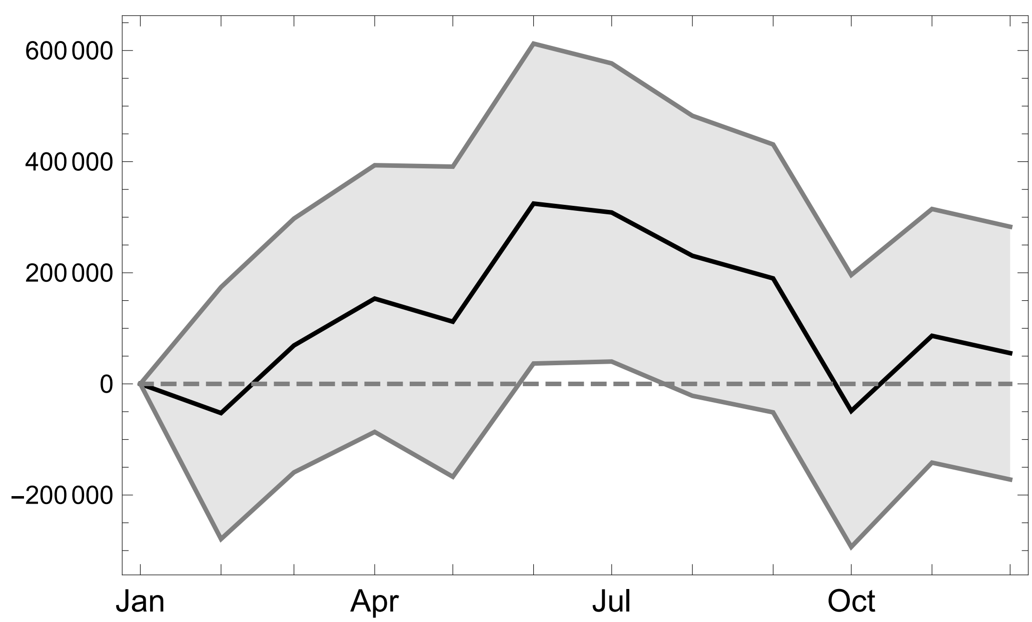
Constant	71298.99
	(80049.09)
Observations	36
R-squared	0.96

+ p<0.10, * p<0.05, ** p<0.01

(OLS regression coefficient with standard errors in parentheses.)

As seen in Table 6, the null hypothesis is rejected since the p-value for n_g is below all the thresholds. This means that refugee arrivals have strong effects on the number of total tweets. For June, July, and August, the p-value lies below one or two of the thresholds. This could mean that in these months, there is an increase in the number of tweets that cannot be attributed to the arrival of refugees. This yearly modulation can be seen in Figure 2 where I show the values of the coefficients b_m for all the months, together with the 95% confidence level intervals. It can be deduced that the coefficients are highly compatible with being zero.

Figure 2
Yearly Modulation for Number of Tweets



Coefficients b_m (black) and 95% confidence level intervals (grey) for all the years in the model for number of tweets.

Correlation does not always equal causation. There could be other unknown causes that are not included in this model such as weather and temperature, which could explain both the arrival of refugees and twitter activity. It is worth investigating media coverage of refugees and the Roma during those three years.

3.2.3 Sentiment of Tweets

Here I present the results and analysis of the studies of the correlation between refugee arrivals and average sentiment of tweets about the Roma in the thirty largest cities in Italy. In table 7, I show the values of the parameters of the liner model considered above.

Table 7

Sentiment of Tweets in Time Series Explained

Parameter	Value
Refugees	4.74e-06**
	(8.26e-07)
January	0.00
	(.)
February	-.0028004
	(.0194745)
March	-.0017311
	(.0196219)
April	.0073844
	(.0206181)
May	.0129647
	(.0239757)
June	.0157529
	(.024733)
July	.0090798
	(.0230499)
August	-.0111996
	(.0216491)
September	-.0109128
	(.0207145)
October	-.0224583

	(.0210231)
November	-.0144836
	(.0196086)
December	-.014578
	(.0195438)
Constant	.6523908**
	(.0142272)
Observations	36
R-squared	0.84

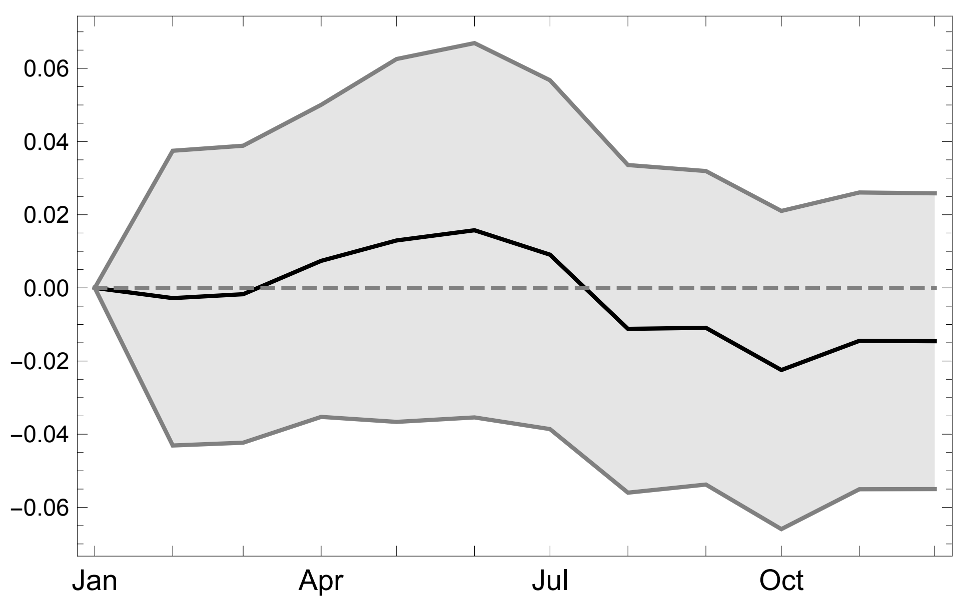
+ p<0.10, * p<0.05, ** p<0.01

(OLS regression coefficient with standard errors in parentheses.)

As seen in Table 7, the null hypothesis is rejected since the p-value for s_g is below all the thresholds. In this case, all the p-values for parameters b_m are high enough to consider them highly compatible with zero. The effect of the yearly modulations is then much smaller in the case of sentiment. The value of these coefficients can be seen in Figure 3 below, together with the 95% confidence level intervals.

Figure 3

Yearly Modulation for Sentiment of Tweets



Coefficients b_m (black) and 95% confidence level intervals (grey) for all the years in the model for sentiment of tweets.

3.2.4 Auto-correlation Study

In order to confirm the validity of the presented results in times series analysis above, I intend to analyze the data looking for auto-correlation. If the number of tweets data shows a high level of auto-correlation, it would mean that high level of Twitter activity at some time would be the main cause for high level of Twitter activity at a later time. In this scenario, the influence of refugee arrivals on the number of tweets could disappear. This applies to the average sentiment as well. In order to do this, I add to the linear model the lagged dependent variables as independent variables. Since monthly data is used in the model, the lag is a month.

3.2.4.1 Number of Tweets

Here I present the auto-correlation analysis of the number of tweets.

Table 8

Number of Tweets in Time Series Explained (Auto-correlation)

Parameter	Value
Refugees	49.30**
	(5.27)
January	0.00
	(.)
February	-11782.95
	(124270.1)
March	115286.1
	(125663.2)
April	188105.6
	(128502.7)
May	129712.9
	(143361.2)
June	285969.8*
	(147837)
July	211559.8
	(149041)

August	149435.6
	(139029.9)
September	135734.7
	(129918.4)
October	-69695.69
	(128275.4)
November	65133.85
	(121688.6)
December	63508.97
	(121383.8)
Lagged N[10]	.1484097
	(.0921045)
Constant	8843.877
	(105231.4)
Observations	35
R-squared	0.96

+ $p<0.10$, * $p<0.05$, ** $p<0.01$

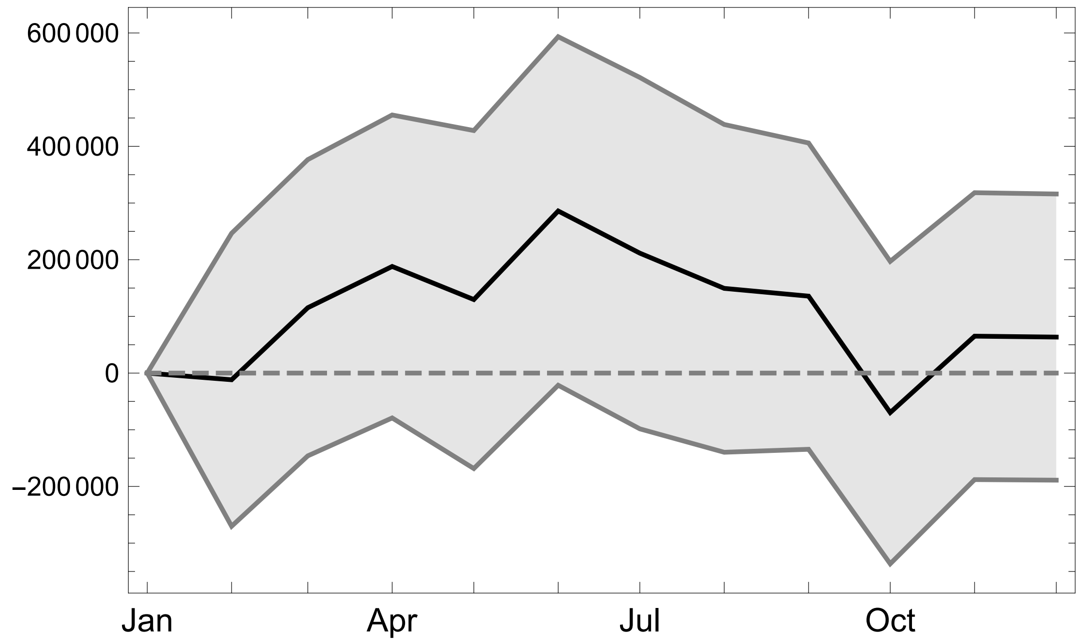
(OLS regression coefficient with standard errors in parentheses.)

As can be seen in Table 8, the correlation with refugee arrivals is still highly significant. The coefficient for refugee arrivals went down by a mere 4.2. Also, one can see in Figure 4 that, in general, the monthly contribution becomes less significant. Overall, however, the effect of the lagged dependent variable is minor, and the model is highly robust.

[10] Lagged number of tweets.

Figure 4

Yearly Modulation for Number of Tweets (auto-correlation)



Coefficients b_m (black) and 95% confidence level intervals (grey) for all the years in the model for number of tweets (auto-correlation).

3.2.4.2 Sentiment of Tweets

Here I present the auto-correlation analysis of the average sentiment.

Table 9

Sentiment of Tweets in Time Series Explained (Auto-correlation)

Parameter	Value
Refugees	3.55e-06**
	(6.43e-07)
January	0.00
	(.)
February	.015219
	(.015803)
March	.0167443
	(.0158949)
April	.0265839
	(.0165105)
May	.0246573

	(.0184493)
June	.0043455
	(.0192048)
July	-.0118336
	(.0186059)
August	-.0237617
	(.0173022)
September	-.0073154
	(.0163201)
October	-.0105244
	(.0165658)
November	-.006827
	(.015661)
December	.0047204
	(.0158829)
Lagged S[11]	.5336722**
	(.1108317)
Constant	.2818082**
	(.0768201)
Observations	35
R-squared	0.92

+ p<0.10, * p<0.05, ** p<0.01

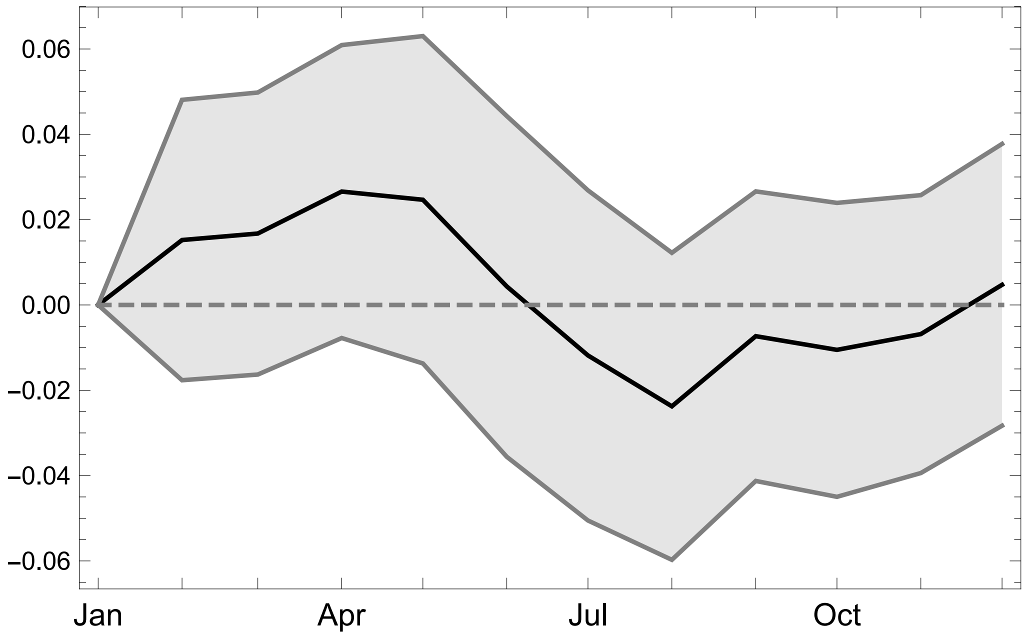
(OLS regression coefficient with standard errors in parentheses.)

As can be seen in table 9, the coefficient of refugee arrivals went down by about 25% even though it is still significantly non-zero. Here I find that the auto-correlation is significant as well since its p-value passes all the tests. Sentiment towards the Roma appears to have some of its own inherent dynamics that are independent of the other predictors. Nevertheless, this does not destroy the correlation with refugee arrivals nor changes the monthly contributions significantly as can be seen in Figure 5 below.

[11] Lagged sentiment of tweets.

Figure 5

Yearly Modulation for Sentiment of Tweets (auto-correlation)



Coefficients b_m (black) and 95% confidence level intervals (grey) for all the years in the model for sentiment of tweets (auto-correlation).

Conclusion

Institutionalized racism is justified by both local and national governments in Italy. The dismantling of illegal Roma camps proposed by mayors in big cities (notably Rome, Milan, and Florence) and the requirement for the Roma to have their finger prints collected are just a few of the examples (The European Roma Rights Centre 2018). However, my analysis showed that negative attitudes towards the Roma are to a good degree determined by factors that have nothing to do with the Roma themselves. When immigrants who are not Roma threaten the physical and economic safety of Italians, the Roma seem to get mixed up with other immigrant groups. The Roma can be a perfect target for hate speech online when Italians are frustrated by an increasing number of immigrants and refugees from Africa, Eastern Europe and the Middle East.

For this research, I did not pay attention to the level of news coverage of newly arrived refugees in Italy and how Italians think of them in general. The time series analysis result was a "long shot" as I see no direct connection between refugees and the Roma. However, the data suggests that many Italians may be prone to group racial and ethnic minority groups as "others," and talk negatively about the Roma or other minority groups when they feel threatened by a minority group of refugees.

The methods I used for this study can be applied to study people's attitude towards other minority groups in other countries. How people's fear of and hate towards a minority group can be exacerbated by the presence of something that is not directly related to them can be tested by using the same methods. A minority group does not have to be a racial or ethnic one. For example, sexual minorities and religious minorities can be included as well. Although my research design has many flaws, I hope this serves as a starting point for other researchers and myself to continue exploring the process of otherization using social media data.

Opportunities for Future Research

It is a limitation of my study that my analysis is only based on Twitter data over the span of three years, which may raise issues of generalizability. Although there is literature that discovered a relatively strong correlation between public opinion and what people write on Twitter, the relationship is not concrete. Also, we need to keep in mind the demography of Twitter users. According to Pew Report of Social Media Usage (2018), only 13% of those aged fifty to sixty-four, and only six percent of those aged over sixty-five use Twitter. Facebook posts tend to be private, which makes it difficult to mine data without paying the company. Instagram's focus is on photography, rather than short texts that often accompany a photograph. Not so many people bother to create a blog and write a long opinion about current social or political issues. Therefore, despite those limitations, Twitter is still a useful platform for measuring public opinion.

It will be useful to keep collecting data and pay attention to political events and immigration trends to see how those things are influencing people's behavior on Twitter. In measuring sentiment of tweets, social scientist will benefit from drawing up clear guidelines and having software developed by computer engineers specifically for sentiment analysis of microblogging platforms. Mining and managing data is a time and cost consuming process.

The linear models used in this study would not work well if some of the conditions here were not met. For example, if the group under consideration became large within the country, it is clear that their ideas would become relevant on social media. The linear models for sentiment has intrinsic inconsistencies. If the independent variables increased enough, the sentiment would take values above one, which are meaningless. More sophisticated models should be studied.

Bibliography

Achim, Vlorel. 1998. *The Roma in Romanian History.* Budapest: Central European University Press.

Aragona, Eleonora. 2015. "La Mappa Dei Rom in Italia. La Comunità Conta 150/170 Mila Persone." *Info Data*. Retrieved December 22, 2016. (http://www.infodata.ilsole24ore.com/2015/09/29/la-mappa-dei-rom-in-italia-la-comunita-conta-150170-mila-persone/).

Aslam, Salman. 2018. "Twitter by the Numbers: Stats, Demographics & Fun Facts." *Omnicore.* Retrieved March 1, 2018. (https://www.omnicoreagency.com/twitter-statistics/).

Associazione 21 Luglio. 2015. *Rapporto Annuale.* Rome: Associazione 21 Luglio. Retrieved December 22, 2016. (http://www.21luglio.org/wp-content/uploads/2015/04/Rapporto-annuale-Associazione-21-luglio.pdf).

Balasubramanyan, Ramnath, Brendan O'Connor, Bryan R. Routledge, and Noah A. Smith. 2010. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* 2010:122-129.

BBC News. 2007. "Italy Starts Deporting Romanians." *BBC News.* Retrieved December 10, 2016. (http://news.bbc.co.uk/2/hi/europe/7078532.stm).

Bollen, Johan, Huina Mao, and Alberto Pepe. 2011. "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena." *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* 2011:450-453.

Bosker, Bianca. 2012. “9 Quirkiest Facts About Twitter: Gaze Into The Soul Of The Twittersphere.” *Huffington Post.* Retrieved December 10, 2016.

Brownlee, Jason. 2017. “How to Develop a Deep Learning Bag-of-Words Model for Predicting Movie Review Sentiment.” *Machine Learning Mastery.* Retrieved December 10, 2017. (https://machinelearningmastery.com/deep-learning-bag-of-words-model-sentiment-analysis/).

Burnap, Pete, and Matthew L. Williams. 2015. "Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making." *Policy & Internet* 7.2: 223-42.

Carrigan, Mark. 2014. “The Six Challenges Social Media Poses for Social Researchers.” *Marc Carrigan.* Retrieved December 9, 2016. (https://markcarrigan.net/2014/05/21/the-six-challenges-social-media-poses-for-social-researchers/).

Ciobanu, Claudia. 2008. “Europe: Home to Roma, and No Place for Them.” *Inter Press Service.* Retrieved December 10, 2016. (http://www.ipsnews.net/2008/05/europe-home-to-roma-and-no-place-for-them/).

Colombo, Asher and Giuseppe Sciortino. 2001. “The Flows and the Flood: The Public Discourse on Immigration in Italy, 1969-2001.” *Journal of Modern Italian Studies*, 9.2:194–113.

DeMers, Jayson. 2017. “How is Twitter Still Alive?”. *Forbes.* Retrieved December 10, 2017. (https://www.forbes.com/sites/jaysondemers/2017/03/09/4-reasons-twitter-will-die-and-5-reasons-it-wont/#f27586a39d8b).

Hancock, Ian. 1997. “A Glossary of Romani Terms.” *The American Journal of Comparative Law,* 45.2:329-344.

Hancokc, Ian and Hristo Kyuchukov. 2005. *A History of the Romani People.* Honesdale: Boyds Mills Press.

Hancock, Ian. 2002. *We Are the Romani People.* Hatfield: University of Hertfordshire Press.

Larson, Selena. 2017. “Welcome to a World With 280-Character Tweets.” *CNN.* Retrieved December 10, 2017. (http://money.cnn.com/2017/11/07/technology/twitter-280-character-limit/index.html).

Leask, David. 2016. “Inside the Squalor of Italy's Roma camps.” *The Herald Scotland.* Retrieved December 11, 2017. (http://www.heraldscotland.com/news/14594904.Inside_the_squalor_of_Italy_s_Roma_camps/).

Matras, Yaron. 2015. *The Romani Gypsies.* New York: Harvard University Press.

Mitchell, Travis. 2016. *Negative Opinions about Roma, Muslims in Several European Nations - Pew Research Center's Global Attitudes Project -*. Washington, DC: Pew Research Center. Retrieved December 22, 2016. (http://www.pewglobal.org/2016/07/11/europeans-fear-wave-of-refugees-will-mean-more-terrorism-fewer-jobs/lede-chart-2/).

Orta, Lucy. 2010. *Mapping the Invisible: Eu-Roma Gypsies*. London: Black Dog Pub.

Pew Research Center. 2014. *Views of Roma, Muslims, Jews - Pew Research Center's Global Attitudes Project -*. Washington, DC: Pew Research Center. Retrieved

December 22, 2016. (http://www.pewglobal.org/2014/05/12/chapter-4-views-of-roma-muslims-jews/)

Picker, Giovanni. 2011. "Welcome 'in'. Left-wing Tuscany and Romani Migrants (1987-2007)." *Journal of Modern Italian Studies* 16.5: 607-20.

Roma Europe. 2012. *Roma Migrants from Bulgaria and Romania - Migration Patterns and Integration in Italy and Spain.* Paris: Roma Europe. Retrieved January 10, 2018. (https://www.romeurope.org/IMG/pdf/Migration_fin.pdf).

Russo, Famcesco. 2017. "Social Media in Italy: 2016 Trends." *Zucchetti Centro Sistemi.* Retrieved December 10, 2017. (https://www.zcscompany.com/en/blog/post/social-media-in-italia-nel-2016-ecco-qualche-dato-aggiornato).

Sigona, Nando. "How Can a `Nomad' Be a `Refugee'?" 2003. *Sociology*, 37.1: 69–79.

Sigona, Nando. 2011."The Governance of Romani People in Italy: Discourse, Policy and Practice." *Journal of Modern Italian Studies* 16.5: 590–606.

Stieglitz, Stefan, and Linh Dang-Xuan. 2013. "Emotions and Information Diffusion in Social Media Sentiment of Microblogs and Sharing Behavior." *Journal of Management Information Systems* 29.4: 217-48.

The European Union Agency for Fundamental Rights. 2011. *Results from the 201 Roma Survey.* Brussels: European Union Agency for Fundamental Rights. Retrieved December 11, 2017. (http://fra.europa.eu/en/publications-and-resources/data-and-maps/survey-data-explorer-results-2011-roma-survey).

The European Union Agency for Fundamental Rights. 2017. *Second European Union Minorities and Discrimination Survey - Main results.* European Union Agency for

Fundamental Rights. Retrieved December 11, 2017. (http://fra.europa.eu/en/publication/2017/eumidis-ii-main-results).

The Italian National Institute of Statistics. 2018. *Demography in Figures.* Rome: Italian National Institute of Statistics. Retrieved January 22, 2018. (http://demo.istat.it/index_e.html).

The Italian National Institute of Statistics. 2014. *I Cittadini Non Comunitari Regolarmente Soggiornanti.* Rome: Italian National Institute of Statistics. Retrieved December 22, 2016. (http://www.istat.it/it/archivio/129854).

The Telegraph. 2011. "Holocaust Memorial Day: 'Forgotten Holocaust' of Roma finally acknowledged in Germany." *The Telegraph.* Retrieved December 10, 2016. (https://www.telegraph.co.uk/history/world-war-two/8284274/Holocaust-Memorial-Day-Forgotten-Holocaust-of-Roma-finally-acknowledged-in-Germany.html).

The United Nations High Commissioner for Refugees. 2018. *Mediterranean Situation*. Geneva: United Nations High Commissioner for Refugees. Retrieved January 15, 2018. (https://data2.unhcr.org/en/situations/mediterranean/location/5205).

(https://www.ushmm.org/wlc/en/article.php?ModuleId=10005219).

The United States Holocaust Memorial Museum. "Genocide of European Roma (Gypsies)." *Holocaust Encyclopedia.* Retrieved December 10, 2016. (https://www.ushmm.org/wlc/en/article.php?ModuleId=10005219).

Twitter. "Our Values." *About Twitter.* Retrieved December 10, 2016. (https://about.twitter.com/content/about-twitter/en_us/values.html).

USC Shoah Foundation. 2013. "The Roma and Sinti People in Italy." *Giving Memory a Future the Sinti and the Roma in Italy and around the World."* Retrieved December

10, 2016. (https://sfi.usc.edu/education/roma-sinti/en/conosciamo-i-roma-e-i-sinti/chi-sono/nel-mondo-e-in-italia/i-rom-e-i-sinti-in-italia.php).

Woodfield, Kandy. 2014. "Future Challenges, Take-down Notices & Social Media Research." *#NSMNSS New Social Media, New Social Science?* Retrieved December 10, 2016. (http://nsmnss.blogspot.co.uk/2014/05/future-challenges-take-down-notices.html).