The Islamic University–Gaza
Research and Postgraduate Affairs
Faculty of Information Technology
Master of Information Technology

الجامعــــــــة الإسـلاميـة – غـزة
شئون البحث العلمي والدراسات العليا
كـليـــة تكنـولـوجيـا المعلــومـــات
ماجستير تكنـولـوجيـا المعلــومـــات

An Analytical Study of Perception of Palestine in Western Social Media Based on Twitter

دراسة تحليليه لتوجهات وسائل التواصل الاجتماعي الغربية حول فلسطين باستخدام تويتر

Osama Abdullah AbuDhrooj

Supervised by:

Dr. Iyad M. Alagha

Assistant professor of Computer sciences

**A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Information Technology**

August/2017

إقـــــــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

An Analytical Study of Perception of Palestine in Western Social Media Based on Twitter

دراسة تحليليه لتوجهات وسائل التواصل الاجتماعي الغربية حول فلسطين باستخدام تويتر

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وأن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل الاخرين لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

Declaration

I understand the nature of plagiarism, and I am aware of the University’s policy on this.

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted by others elsewhere for any other degree or qualification.

Student’s Name:	أسامة عبد الله أبو دحروج	اسم الطالب:
Signatures:	أسامة أبو دحروج	التوقيع:
Date:	18/9/2017	التاريخ:

عمادة البحث العلمي والدراسات العليا **هاتف داخلي:** 1150

الرقم: Ref: ج س غ/35/

التاريخ: Date: 2017/08/30

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة عمادة البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحث/ **أسامة عبد الله حسن أبو دحروج** لنيل درجة الماجستير في **كلية *تكنولوجيا المعلومات* برنامج تكنولوجيا المعلومات** وموضوعها:

دراسة تحليلية لوجهات النظر حول فلسطين في وسائل التواصل الاجتماعي الغربية بالإعتماد على . توتير

An Analytical Study of Perception of Palestine in Western Social Media Based on Twitter

وبعد المناقشة التي تمت اليوم الأربعاء 08 ذو الحجة 1438هـ، الموافق 2017/8/30م الساعة الحادية عشر صباحاً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

د. إيـــاد محمـــد الأغـــا	**مشـرفاً و رئيسـاً**	الآغا ....
د. راويـة فـوزي عـوض الله	**مناقشـاً داخليـاً**	....
د. تـــامر ســـعد فطـــاير	**مناقشـاً خارجيـاً**	....

وبعد المداولة أوصت اللجنة بمنح الباحث درجة الماجستير في **كليـة *تكنولوجيـا المعلومـات*/ برنـامج تكنولوجيا المعلومات.**

واللجنة إذ تمنحه هذه الدرجة فإنها توصيه بتقوى الله ولزوم طاعته وأن يسخر علمه في خدمة دينه ووطنه.

والله ولي التوفيق ،،،

عميد البحث العلمي والدراسات العليا

أ.د. مازن اسماعيل هنية

Abstract

Sentiment analysis is the field of study that analyses people's opinions towards products, events, individuals, and their attributes. The international public opinion with respect to Palestine is still divided and vague. There are few official opinion polls that were carried out to measure the perception towards Palestine. Besides, opinion polls have various and well-known limitations such as the small sample size and the difficulty to perform fine-grained analysis of poll results. Alternatively, this research proposes an analytical study to measure the public's opinion on the Palestinian Issue by analysing Twitter data.

A dataset consisting of hundreds of thousands of tweets were collected from Twitter. First, sentiment analysis will be performed on the collected tweets. Second, polarities were analysed at two different levels: country-level and individual level: The country-level analysis aims to explore the country's overall interest in and attitude towards Palestine by: 1) Identifying counties that generate the most Palestine-focused tweets, 2) Measuring the friendliness of each country towards Palestine. 3) Analysing time series data to investigate the changes of attitude over time.

The individual-level analysis aims to analyse data based on the activity of individuals. The attitudes of both opinion leaders and Arab ethnicities were analysed and discussed in light of the counties' attitudes. Results have shown that the superpower countries generate the most Palestine-focused tweets but they have less favourable views of Palestine. Despite the low friendliness scores reported for most countries, high levels of divergence in public opinion were observed, indicating the increasing influence of the pro-Palestinian sectors over the globe. Furthermore, opinion leaders show more friendly attitudes towards Palestine when compared to the overall country's attitude. Results also showed that Arab users have more positive attitude towards Palestine than non-Arabs. However, they have not caused significant changes to the overall attitudes of their countries.

Keywords: *Palestine, Perceptions, Attitudes, Tweets, Sentiment Analysis*

الملخص

تحليل المشاعر هو عبارة عن مجال تحليل أراء الناس تجاة منتجات أو أحداث أو أشخاص والخصائص المتعلقة بهم. لا يزال الرأي العام الدولي فيما يتعلق بفلسطين منقسما ومبهما. هناك عدد قليل من استطلاعات الرأي الرسمية التي أجريت لقياس الإدراك تجاه فلسطين. وبالإضافة إلى ذلك، فإن استطلاعات الرأي لديها قيود متعددة ومعروفة مثل حجم العينة الصغيرة وصعوبة إجراء تحليل دقيق الحبيبات (دقيق التعريق) لنتائج الاستطلاع. و كبديل، يعرض هذا البحث دراسة تحليلية لقياس رأي الجمهور حول القضية الفلسطينية من خلال تحليل بيانات توتير.

تم تجميع مجموعة بيانات تتكون من مئات الآلاف من التغريدات المأخوذة من تويتر. أولا، سيتم اجراء تحليل المشاعر على التغريدات التي تم جمعها. ثانيا، تم تحليل الاستقطاب على مستويين مختلفين: المستوى الدولي والفردي: يهدف التحليل على المستوى الدولي إلى استكشاف الاهتمام العام للبلد وموقفه تجاه فلسطين من خلال: 1) تحديد الدول التي تولد معظم التغريدات التي تركز على فلسطين، 2) قياس مدى الود لكل دولة نحو فلسطين.3) تحليل بيانات السلسلة الزمنية للتحقيق في تغييرات الموقف مع مرور الوقت.

ويهدف التحليل على مستوى الفرد إلى تحليل البيانات على أساس نشاط الأفراد. وتم تحليل مواقف قادة الرأي والعرقيات العربية ومناقشتها في ضوء مواقف الدول. وأظهرت النتائج أن دول القوة العظمى تولد معظم التغريدات التي تركز على فلسطين، ولكن لديها آراء أقل إيجابية من فلسطين. وعلى الرغم من انخفاض درجات الود في معظم البلدان، لوحظت مستويات عالية من التباين في الرأي العام، مما يشير إلى تزايد تأثير القطاعات المؤيدة للفلسطينيين في جميع أنحاء العالم. وعلاوة على ذلك، يظهر قادة الرأي مواقف أكثر ودية تجاه فلسطين بالمقارنة مع موقف البلاد عموما. وأظهرت النتائج أيضاً أن الأفراد العرب لديهم موقف ايجابي أكثر تجاه فلسطين من غير العرب، بيد أنها لم تسبب تغييرات كبيرة في المواقف العامة لبلدانها.

كلمات مفتاحية: فلسطين، التصورات، المواقف، التغريدات، تحليل المشاعر

Dedication

To my dear mother and father who could not see this thesis complete.

To my wonderful family. To my brothers and sister for their love and care

To all my friends and colleagues who supported me.

Acknowledgment

At the very outset, my thankfulness is to Allah the almighty who provided me with the needed strength to successfully accomplish this work, and to be surrounded by great and helpful people.

I would like to express my deepest gratitude to my advisor, **Dr. Iyad Mohammed Al Agha**, for his constant guidance, challenging discussions and advices, enthusiasm, and knowledge. He motivated me to think more deeply about my work. He also made great effort to build the structure and refine every detail of my work. I am grateful to him for working with me. I learned so much, it has been an honor. My Allah reword him on my behalf.

Also, I am thankful for my whole family and friends for encouragement and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API	Application Programming Interface
NLP	Natural Language Processing
POS	Part-of-Speech
HTTP	Hypertext Transfer Protocol

Chapter 1

Introduction

Chapter 1
Introduction

1.1 Introduction

The Palestinian-Israeli conflict is one of the most complicated and intractable conflicts which began around the turn of the 20th century. While the ongoing peace process between the Palestinians and Israelis, there are no signs, unfortunately, of a solution in sight. This is due to the official attitudes of the Western governments that sympathize with Israel all-time. While the public opinion towards Palestinian rights is still different and varying, there are few official opinion polls that were carried out in Western courtiers to measure the perception towards Palestine. But these opinion polls are not always reliable because since they have several limitations:

First, polls often use samples of small sizes; increasing the sample size may result in significant increase in time and effort.

Second, polls often do not measure the influence of daily events that occur in Palestine.

Third, traditional polls will be difficult to perform fine-grained analysis. For example, it will be difficult to determine the impact of ethnicity, leadership, location on the results.

In this age, social-media networks are very popular where people can share attitudes or opinions with other people or towards an opinion target rapidly. The data generated through social media networks has become an important topic for scientific researches.

User' opinions on social media networks may be used for discovery and recognition of fine or bad expression on different subjects of interest. Even though online reviews may be determined using conventional methods, but this way is insufficient thinking about the massive extent of facts generated on social media networks. This reality emphasizes the importance of data mining strategies in mining opinion expressed on many application domains including politics, marketing, education, law, etc. (Osimo & Mureddu 2012). Data mining used for opinion mining include sets of techniques: *Feature-Based*, *Opinion Definition and Opinion Summarization*, and *Opinion Extraction* (Godbole, Srinivasaiah, & Skiena 2007).

Twitter is one of the most common examples of social media networks that allows users to share short information known as tweets which are limited to 140 characters (Lohmann, Burch, Schmauder, & Weiskopf 2012; Sarlan, Nadam, & Basri 2014). Twitter is a perfect platform for the extraction of general public opinion on certain issues (Pak & Paroubek 2010).

Twitter has been used as a data source for successfully predicting climate change (An, Ganguly, Fang, Scyphers, Hunter, & Dy 2014), as well as predicting stock market (Bollen, Mao, & Zeng 2011). In the world of politics there are many works used tweets to predict election result such as (Ceron, Curini, & Iacus 2015; Ibrahim, Abdillah, Wicaksono, & Adriani 2015; Jose & Chooralil 2015; Tumasjan, Sprenger, Sandner, & Welpe 2010).

Sentiment analysis, also called opinion mining, is defined as the technique of determining the emotional tone in the back of a sequence of words, used to gain an knowledge of the attitudes, opinions and emotions expressed in textual content (Liu 2012). Sentiment analysis is the software of Natural Language Processing (NLP) methods to recognizing whether the given textual entity is subjective or objective, and identifying polarity of subjective texts (Roebuck 2011). Mostly, sentiment analysis is particularly beneficial in social media networks checking as it lets us to acquisitions an overview of the extensive public opinion at certain topics. In the realm of politics, there have been many researches that have analyzed the public sentiment found on Twitter (Di Fatta, Reade, Jaworska, & Nanda 2015; Montesinos, Rodr, Orchard, & Eyheramendy 2015; Zhou, Tao, Yong, & Yang 2013).

In our work, we propose an approach to analyze and measure the public perception towards Palestine by using data from Twitter. A dataset consisting of hundreds of thousands of tweets will be collected from Twitter by using appropriate Application Programming Interface (API). First, sentiment analysis will be performed on the collected tweets to measure the polarity. Second, several features will be extracted from the tweets including features related to leadership, ethnicity, country, etc. These features were chosen because they include a large fragment of the country and also have an impact on decision-making. Relationships between the polarity scores obtained from sentiment analysis and the extracted features will be identified. Due to the massive size of the dataset, a parallel computation framework will be used to pre-process and analyze the dataset. We need to use Apache Spark cluster (Karau, Konwinski, Wendell, & Zaharia 2015) to perform our analysis. Figure (1.1) presents the approach as simple steps, details will be discussed later in Chapter 3 and Chapter 4.

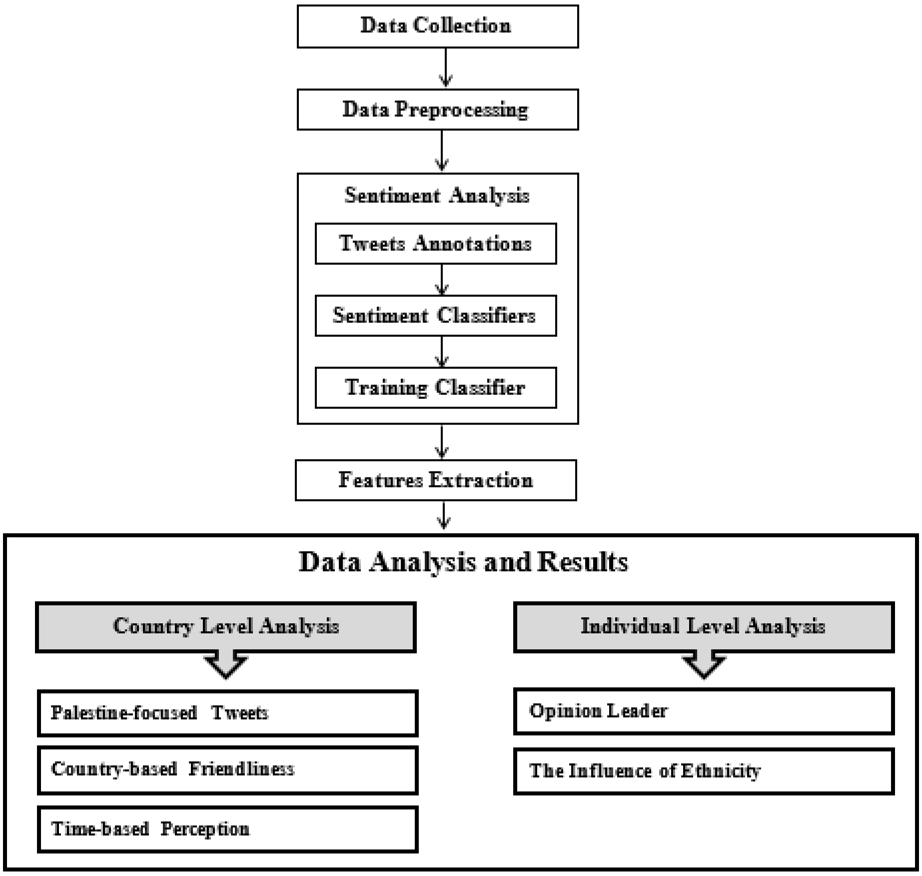

Figure (1.1): Steps of our approach

1.2 Statement of the problem

The perception of Palestine in Western media is still divided and vague. There are few official and trustworthy opinion polls that were carried out to measure the perception towards Palestine. Besides, opinion polls have various and well-known limitations such as the small sample size, the inability to measure the influence of daily events, and the difficulty to perform fine-grained analysis of poll results. This research proposes an analytical study to measure the public's opinion on the Palestinian Issue by analysing Twitter data.

1.3 Objectives

In this section, we present both the main and the specific objectives of the research work.

1.3.1 Main Objective

Propose an approach to analyze and measure the public perception towards Palestine by using data from Twitter. This will be achieved by performing sentiment analysis on the collected tweets, and then link the polarity scores with other features extracted from tweets. The proposed approach will focus on the major Western native English-speaking counties.

1.3.2 Specific Objectives

The proposed approach will try to explore the attitudes towards Palestine through the following:

- Analysis of the attitudes towards Palestine at the country level.
- Analysis of the attitude towards Palestine to investigate how the country's sentiment changes over time, and the rationale behind these changes.
- Analysis of the country leader's opinions.
- Analyse the influence of ethnicity on the public opinion, and the potential relation between the ethnicity of user and its attitude towards Palestine.

1.4 Importance of Research

This study will be a significant endeavour in clarifying the role of Twitter in changing the public opinion towards Palestinian rights. This study will also be beneficial to the researchers in the field of study in different concepts related to the use of twitter in such issues. The study will offer insight into the perception of Palestine by analysing aspects that have not been covered in traditional polls. Results of the study may be also beneficial to politicians and policy makers who are interested in understanding the attitude towards Palestine in social media.

1.5 Scope and limitations of the project

- The analysis will be restricted to some major Western English-speaking countries. Other languages will be out of scope.

- A considerable amount of Twitter data is not associated with location information. Data with unidentified locations will be excluded.
- The study of ethnicity effect will be limited to Arab population residing in the Westerns counties. Other ethnicities will not be studied.
- Sentiment analysis will be carried out using open source tools such as Stanford CoreNLP (Manning, Surdeanu, Bauer, Finkel, Bethard, & McClosky 2014). We will not implement our own sentiment analysis tool. However, several tools will be compared to choose the best one.

1.6 Research contribution

The work in this thesis has the following research contributions:

1. Identifying counties that generate the most Palestine-focused tweets by aggregating tweets at the county level.
2. Measuring the friendliness of each country towards Palestine.
3. Analysing time series data to investigate how the country's attitude towards Palestine changes over time.
4. Capturing the attitudes of opinion leaders, and measuring their friendliness towards Palestine.
5. Analysing the influence of ethnicity on the public opinion, and the potential relation between the ethnicity of users and their attitudes towards Palestine.

1.7 Structure of Thesis

The thesis consists of five chapters. The chapters are organized in general as follows:

Chapter 1: Introduction: this chapter is an overview of the problem, work done in the field, and focuses on the proposed solution.

Chapter 2: Literature Review: this chapter focuses on related works that employed sentiment analysis and polarity classification.

Chapter 3: Methodology: This chapter explains the detailed steps of the sentiment analysis.

Chapter 4: Data Analysis and Results: this chapter extracts features that need to analyse country and individuals' perceptions.

Chapter 5: Conclusions: this chapter presents a conclusion of the thesis and possible future works.

Chapter 2

Literature Review

Chapter 2
Literature Review

2.1 Introduction

Using social media to expect the opinion mining and sentiment analysis become very important in latest years. Sentiment analysis has been handled as a Natural Language Processing task at many levels such as a document level classification task (Pang, Lee, & Vaithyanathan 2002; Turney 2002), sentence level classification (J. M. Wiebe, Bruce, & O'Hara 1999), and entity level classification (Hu & Liu 2004).

In the following sections, we present a brief background about Sentiment analysis and opinion mining, Apache Spark, Moore and McCabe's Method, and Twitter insight.

2.2 Background

2.2.1 Sentiment Analysis and Opinion Mining

Sentiment analysis, also known as opinion mining, is the field of study that analyses people’s opinions, attitudes, and emotions towards products, events, individuals, and their attributes. Sentiment analysis and opinion mining primarily emphasis on opinion which express or involve positive or negative sentiments. Actually, sentiment analysis is now right at the core of social media studies. Hence, studies in sentiment analysis have an influence on political, economics, and social science as they all changed by people’s opinions.

2.2.1.1 Different Level of Analysis

In general, sentiment analysis has been examined principally at three level by the following.

- **Document level:** This task is usually known as document level sentiment classification. The mission at this level is to decide whether an entire opinion document expresses as a positive or negative sentiment (Pang et al. 2002); (Turney 2002). For instance, given a laptop posts, the system determines whether the posts express an overall positive or negative opinion about the laptop.

- **Sentence level:** The task at this level focus on sentence and goes to decide whether each sentence expressed a positive, negative, or neutral opinion. This type of analysis is closely related to subjectivity classification (J. M. Wiebe et al. 1999), which determine sentences (known objective sentences) that express actual information from sentence (known subjective sentences) that express subjective views and opinions.

- **Entity and Aspect level:** Aspect level executes fine-grained analysis. This level was previously called feature-based opinion mining and summarization (Hu et al. 2004) .Instead of looking at documents or sentences, aspect level directly looks at the opinion itself. It is based on the concept that an opinion contains of sentiment (positive or negative) and aim of opinion. consequently, the main goal of this level is to detect sentiments on entities and/or their aspects.

2.2.1.2 Sentiment Lexicons and Its problems

The most significant pointers of sentiments are sentiment words, which known opinion words. These words are usually used to express positive or negative sentiments. For instance, good, fantastic, and surprising are positive sentiment words, and evil, poor, and horrible are negative sentiment words. Regardless of individuals words, there also phrases and idioms. A list of this words and phrases is called a sentiment lexicon. Although, the sentiment lexicon is essential for sentiment analysis but using them is not sufficient for several reasons:

1. A positive or negative word may have inverse orientations in various domains.
2. A sentence holding sentiment words may not express any sentiment. Question sentences and conditional sentences are two important kinds.
3. Sarcastic sentences with or without sentiment words are difficult to deal with.
4. Several sentences without sentiment words can also involve opinions.

2.2.2 The Problem of Sentiment Analysis

There are two significant concepts that are closely related to sentiment and opinion, subjectivity and emotion.

- **Sentence subjectivity**: which can be sub-divided into two types. A subjective sentence expresses some personal opinions, views, or attitudes, while an objective sentence donates some real information about the entity. For instance, “*BMW is a Germany product*” is an objective sentence, while “*I like*

BMW” is a subjective sentence. We can see subjective expression in various forms, e.g., opinions, evaluations, needs, etc. (Riloff, Patwardhan, & Wiebe 2006); (J. Wiebe 2000). The task of determining whether a sentence is subjective or objective is called subjectivity classification (J. Wiebe & Riloff 2005).

- **Emotion:** it defined as our subjective feelings and views. Emotions have been studied in various domains, e.g., sociology and psychology. Emotions are closely related to sentiments. The strength of a sentiment is normally related to the intensity of particular emotions, e.g., happiness and angry.

Based on the above, we notice that the concepts of emotions and sentiment are obviously not equivalent. Opinions express no emotions in a sentence, e.g., *“The light on the laptop is clear”*, and emotions may not have aims, e.g., *“I am so happy today”*

2.2.3 Document Sentiment Classification

Sentiment classification aims to classify an opinion document as expressing a positive or negative opinion. Because this task treats with the whole document as a basic information part, the task is known as the *document-level sentiment classification*. Most existing techniques for document-level classification use supervised learning or unsupervised methods.

2.2.3.1 Supervised Learning

Sentiment classification is generally formulated as a two-class, positive and negative. Training and testing data are used. Sentiment classification is basically a text classification problem. Accordingly, any existing supervised learning method can be used, e.g., naïve Bayes classification, and support vector machines (SVM) (Cristianini & Shawe-Taylor 2000). (Pang et al. 2002) was the initial study to take this technique to categorize movies reviews into positive and negative. The author using unigrams as features in classification performed with either naïve Bayes or SVM. The key for sentiment classification is the selecting of a set of efficient features. Some of these features are as follows.

- ***Terms and their frequency:*** These features are individual words (unigram) and their n-grams with associated frequency counts.

- ***Part of speech:*** The part-of-speech (POS) of each word can be essential too. Words of different parts of speech (POS) may be handled differently.
- ***Sentiment words and phrases:*** Sentiment words are words in a language that are used to express positive or negative sentiments. Separately from individual words, there are also sentiment phrases and idioms.

2.2.3.2 Unsupervised Learning

While sentiment words are frequently the governing element for sentiment classification, it is not hard to assume that sentiment words may be used for sentiment classification in an unsupervised learning. (Turney 2002) performs this technique to classification based on various part-of-speech (POS) tags. Another unsupervised approach is the lexicon-based method, which uses a dictionary of sentiment words and phrases to compute a sentiment score for each document (Taboada, Brooke, Tofiloski, Voll, & Stede 2011). This method was also used in sentence and aspect-level sentiment classification (Hu et al. 2004); (Ding, Liu, & Yu 2008).

2.2.4 Sentence Classification and Subjectivity

In this section, we start in sentence level to categorize sentiment expressed in every sentence. Though, there is no essential difference among document and sentence-level classification because sentence is simply brief document. Sentence classification can be solved in two steps. The first step is to determine whether a sentence is subjective or objective which is called *subjectivity classification*. The second step is classifying subjective sentences into positive and negative classes.

Subjectivity classification classifies sentence into subjective and objective (J. M. Wiebe et al. 1999). An objective sentence expresses some actual information, while a subjective can express opinions, emotions, beliefs, etc.

Most present algorithms to subjectivity classification are depend on supervised learning. For instance, (J. M. Wiebe et al. 1999) performed subjectivity classification using the naïve Bayes classifier with a set of features such as the existence in the sentence of a pronoun, an adjective, and adverb.

(J. Wiebe et al. 2005) proposed an unsupervised approach for subjectivity classification, which used the subjective expression in a sentence to decide the subjectivity of a sentence. The author used distributional similarity (Lin 1998) to discover similar words. However, words obtained had low precision and high recall. Other studies on subjectivity classification of sentences has been done in Arabic (Abdul-Mageed, Diab, & Korayem 2011) based on separate machine learning algorithms using special features.

When a sentence is categorized as subjective, we need to determine whether it expresses a positive or negative. Supervised learning can be used like that for document-level sentiment classification.

2.2.5 Apache Spark

Apache Spark is an open source big data processing framework built around speed, ease of use, and sophisticated analytics. It was started in 2009 as a research project in the UC Berkeley's AMPLab. After one year the code was open sourced and a license was acquire (Zaharia, Chowdhury, Franklin, Shenker, & Stoica 2010).

In our work, we used Spark because it offers a primitive machine learning library called MLlib (Meng, Bradley, Yavuz, Sparks, Venkataraman, Liu, Freeman, Tsai, Amde, & Owen 2016) which has numerous features:

- Resilient Distributed Datasets (RDDs) is essentially a distributed fault-tolerant vector that can perform operation as in local mode.
- RDDs allow user-defined data partitioning, and the execution engine can exploit this to co-partition RDDs.
- Spark logs the history of operations used to build an RDD, enabling reconstruction of lost partitions upon failures.
- Spark provides a high-level API in Java that can be easily extended.

Many works have used Apache Spark for sentiment analysis and other especially in Twitter (Baltas, Kanavos, & Tsakalidis 2016; Compton, Jurgens, & Allen 2014; Nodarakis, Sioutas, Tsakalidis, & Tzimas 2016) .

2.2.6 Moore and McCabe's Method

An outlier is an observation (or measurement) that is different with respect to other values contained in a given dataset. In literature, different definitions of outlier exist: the most commonly referred are reported in the following:

- "Outliers are those data records that do not follow any pattern in an application" (Chen, Fu, & Tang 2002).
- "An outlier is an observation that lies outside the overall pattern of a distribution" (Moore & McCabe 1989).
- "An outlier in a set of data is an observation or a point that is considerably dissimilar or inconsistent with the remainder of the data" (Ramaswamy, Rastogi, & Shim 2000).

Numerous data mining algorithms try to minimize the effect of outliers for instance on a final model to develop, or to remove them in the data pre-processing stage. However, a data miner should be careful when automatically detecting and eliminating outliers because, if the data are correct, their elimination can cause the loss of important hidden information (Kantardzic & Press 2000). Some data mining applications are focused on outlier detection and they are the important result of a data-analysis (Sane & Ghatol 2006).

2.2.6.1 Interquartile range rule

The interquartile range rule is useful in detecting the presence of outliers. The interquartile range rule is the distance between the first and third quartiles $\boldsymbol{IQR = Q_3 - Q_1}$. It is the spread of the center half of the data. The **1.5 * $\boldsymbol{IQR}$ rule** flags observations more than ***1.5 * IQR*** beyond the quartiles as possible outliers.

2.2.7 Twitter insight

One might say that a microblog is a platform in which individuals share short messages, link to different sites, pictures or videos. Commonly a message on a microblog is written by one individual and viewed by a huge number of individuals, which are called followers. Microblog individuals usually perform frequent updates, providing followers some details of information of attention. Microblogs can handle with many various subjects, some are personal, and others where truly interesting facts.

The primary platforms that offer microblogging services are Facebook[(1)], Google+[(2)], YouTube[(3)] and Twitter[(4)].

The initial publication in the Twitter took place on 16 July 2006. To provide the reader an idea of the significance of this platform, Twitter is the 11th most popular site in the world and eighth in the United States, with an average of nearly 313 million monthly active users. These data can be seen with more details in the Twitter section of the Alexa[(5)] website. As with others microblogging service, users often update their status, but in this case the message is restricted to a length of only 140 characters, which are called tweets. The content of the tweets is differed, from individual information, to others where there are links to images, videos, or website that the individual has considered interesting.

The actual accomplishment of Twitter is not in the large number of listed users, or a great many tweets that are posted every day, but in the interest created in the political and business. Political parties and companies know that their users and followers are on Twitter, and what their feelings are in the social network. For this reason, the last years has seen amazing increased in the existence of these activities in Twitter. This shows that Twitter now represent a large amount of information that should not be disparaged and which must be examined in depth. This thus opens up an extensive variety of possibilities for sentiment analysis, opinion mining, information retrieval and so on.

2.3 Related Works

Using social media to expect the opinion mining and sentiment analysis become very important in latest years. Several studies in the field of sentiment analysis in Twitter have been published in political domain (Ceron et al. 2015; Ibrahim et al. 2015; Jose et al. 2015; Salah, Coenen, & Grossi 2014) and polarity classification (Aisopos, Papadakis, Tserpes, & Varvarigou 2012; Hernández & Sallis 2011). In the following

(1) http://facebook.com

(2) http://plus.google.com

(3) http://youtube.com

(4) http://twitter.com

(5) http://www.alexa.com/siteinfo/twitter.com

sections, we present some of these studies proposed by research works in the context of twitter sentiment analysis, as our research work is highly related to these works.

2.3.1 Polarity Classification

In the works related to the sentiment analysis in texts, a difference is made between studies of texts where we suppose that the text is an opinion and consequently need to calculate its polarity, and before measuring polarity it is required to decide whether the text is subjective or objective. As for the research of polarity in Twitter, most studies suppose that tweets are subjective. One of the primary works on the classification of polarity in tweets was done by (Go, Bhayani, & Huang 2009). They performed a supervised classification experiment on tweets in English. They use the emoticons that frequently appear in tweets to distinguish between positive and negative tweets. The authors using Twitter search API to generate a corpus of tweets with positive emoticons ":)" and tweets with negative emoticons ":(". The authors achieved fine results with the Support Vector Machine (SVM), Naive Bayes and maximum entropy algorithms.

Following (Pak et al. 2010) examined the truth of Twitter for the sentiment analysis. They generated a corpus of positive tweets, negative tweets, and neutral tweets. The authors calculate the frequency distribution of words in tweets on corpus of 300,000 tweets generated, and operated a machine learning process to categorize the polarity of tweets by using three algorithms which are Support Vector Machine (SVM), Naive Bayes, and Conditional Random Fields (CRF). The authors concluded that the better algorithms for analysis of opinions on Twitter is to use the Naive Bayes, and use n-grams and Post-tags as features of tweets.

In (Agarwal, Xie, Vovsha, Rambow, & Passonneau 2011), an experiment was operated on the diverse features to be considered in sentiment analysis on Twitter. The experiment is performed on a training data set of tweets labelled manually. The experiment used various methods of polarity classification. The authors start with simple use of unigram, then partial tree kernels model, then use model containing different linguistic features, and finally combination of proposed models. A common feature used is the polarity of the terms showing in every tweet. The authors used the DAL dictionary (Whissell 1989) to calculate the polarity. After wide testing, the authors decided that both partial tree kernels model and features model improve the results, and features that are highly relevant to sentiment analysis are define the polarity on Twitter.

(Aisopos et al. 2012) observe several troubles notable by sentiment analysis on Twitter like the users don’t care about the correct use of grammar, the use a lot of abbreviations, and use of non-standard lexical terms and syntactic patterns. To overcome these troubles, the authors using two different group of models to represent the tweets. First group is models based on the content of every tweet and second group based on models use context of tweet. The authors concluded that the models based on the content model obtains the best results and performance when using the supervised classification.

(Hernández et al. 2011), suggest an unsupervised method which is latent Dirichlet allocation (LDA) for decreasing features in sentiment analysis. The authors assess their technique with 10,000 tweets, downloaded during March and April 2011, in English on the iPad tablet. They used the vector space model to represent tweets and the TF-IDF metric to weight the terms after cleaning the corpus. The authors don’t perform a polarity classification to compare the execution of the whole data set and the reduced data set. They decided that the reduced model is better than the complete model.

(Davidov, Tsur, & Rappoport 2010), used 50 hashtags and 15 emoticons as sentiments labels to build and train K Nearest Neighbors (KNN) algorithms, which is supervised sentiment classifier. The experiments proved by person experts and the results achieved are very favourable.

2.3.2 Using Twitter in Political opinion mining

The utilization of Twitter as a source of information for making predictions is not constrained to the commercial world, but rather it has additionally been applied to predicting election results. One of the main works distributed on this domain was (O'Connor, Balasubramanyan, Routledge, and Smith 2010), where the authors try to exhibit that Twitter can be utilized as a source of information for reviews. They match the evolution of opinions expressed in tweets on three separate subjects that happened through the years 2008 and 2009 with two measurements normally utilized as a part of in conventional reviews. In deciding the polarity of tweets, they used an approach based on unsupervised learning, using the OpinionFinder linguistic resource. For the investigation of the evolution of the opinions, they produced the idea of the daily opinion score, which is basically the proportion amongst between positive and negative tweets. Based on this score, they built a period arrangement and compare it with conventional measurements of opinion polls. The authors achieve that there is a similarity among what is posted on Twitter and conventional opinion polls. They also notice to additional increase the similarity with opinion polls, it would be important to apply more enhanced NLP methods.

With similar aims, (Tumasjan et al. 2010), examine whether it is possible to use Twitter to determine the political opinion and test if political sentiment on Twitter reflects actual life style sentiments about parties and politicians. As a part of their study, they compare party mentions on Twitter with the outcomes of the 2009 German parliament election. They achieved that the relative number of tweets mentioning a party is a suitable predictor for the number of elects of that party in an election.

In the same research domain, (Bermingham & Smeaton 2011), use the Irish general election 2011 as a case study for examination the ability to display political sentiment through mining of social media. The authors use some of machine learning algorithms and assess the error with respect to both polls and the election results. As a final result, they assume that it is ambiguous whether the use of Twitter is a suitable method for examining public sentiment about political matters.

(Maynard & Funk 2011) propose a methodology for determine political opinion from the UK pre-election period in 2010. The method contains of demonstrating every opinionated tweet as a triplet <Person, Opinion, Political Party>. To create this demonstration, the system should recognize the opinion owner, the opinion and the polarity of opinion. They want to find possible correct names that could denote the opinion owner, and to detect the political party by using the entity recognition system ANNIE (A Nearly-New IE System) (Maynard, Tablan, Cunningham, Ursu, Saggion, Bontcheva, & Wilks 2002). The author used an unsupervised methodology based on lexicon approach for the subjective and polarity classification. The authors concluded that the political opinion is more unstable than an opinion about a commercial product.

(Di Fatta et al. 2015), proposed a method used to collect tweets and performing political sentiment index. They collected about 28,473,893 tweets related to UK politics and General Election 2015 for three months from March to May 2015 by using twitter streaming API. They used the parser Penn Treebank to implement political sentiment index, which is based on evaluate words (adjectives) and assigned score: +1 for positive, -1 for negative and 0 for neutral about each word into the text. Based on the above, they have been noticed how a political sentiment index is suitable to discover the important moments in public events and can be used by other data analytics techniques.

(Jose et al. 2015), implemented a novel approach for sentiment analysis using sentiment lexicons such as SentiWordNet and WordNet along with Word Sense Disambiguation (WSD) in order to find political sentiment from real time tweets during the Delhi elections periods. It compared political sentiment towards two politicians using results of sentiment analysis. This was done by applying WSD and negation handling in order to increase accuracy of sentiment analysis. Negation handling results in 1% improvement in classification accuracy and WSD results in 2.6% improvement in classification accuracy.

(Salah et al. 2014),attempt to discover the most appropriate technique for carrying out sentiment analysis with respect to political discussion to predict the attitude of individuals' speakers. The writers compare the operation of three approaches: classification based, generic lexicon based and domain-specific lexicon based through applying these approaches to the 2086 concatenated speeches for 29 various dialogs obtained from the proceeding of the UK house of Commons. The conducted comparison showed that classification based sentiment mining performed best than the lexicon based sentiment mining. Additionally, using domain-specific lexicons produced the better results than the generic lexicon.

(Ibrahim et al. 2015), proposed an approach for predicting the outcomes of Indonesian Presidential Election using twitter to discover the opportunity of easy-to-gather twitter data to be used as a survey supporting tool to recognize public opinion. They implemented a fine-grained political sentiment analysis to separation every tweet into numerous tweets and then allocated every tweet with one of the candidates and its sentiment polarity. Writers recommended that Twitter can perform as a significant resource for any political activity, exactly for expecting the ultimate results of the election itself.

(Zhou et al. 2013), emphasis on users who express their political opinions on Australian federal election 2010 event and suggested a Tweets Sentiment Analysis

Model (TSAM) which can spot the societal interest and common people's reviews in regard to a social event. The experimental outcomes show the efficiency of the system.

(Ceron et al. 2015), apply the supervised method to examine the voting purpose of Twitter users in the United States presidential election 2012. The technique they used offers two significant benefits compared to conventionally employed alternatives: 1) A better understanding to the texts and extra reliable aggregate results. 2) The analysis displays an amazing efficiency of Twitter to ‘‘nowcast’’ as well as to forecast electoral outcomes.

(Younus, Qureshi, Asar, Azam, Saeed, & Touheed 2011), take up an examination of social media through engaging types of users from the developing world through a study of Twitter's role during the recent Tunisian uprising. They used technique for subjectivity classification of tweets matching to political events in the developing world. The writers showed through the experimental evaluations the accuracy of the method was 83.3%, which displays a promising outcome for subjectivity analysis technique.

(Montesinos et al. 2015), proposed work that examined the public's opinion on the presidential primaries for the Alliance political party between Andres Allamand "Renovaci´on Nacional" (RN) and Pablo Longueira from "Uni´on Dem´ocrata Independiente" (UDI) using data gathered from Twitter in the state of Chile. The authors used sentiment analysis to expect the result of the primaries through suggested a dictionary algorithm, which contained of specific positive and negative words, helps in these predictions. The result shows that there is an error rate of 2% compared with the corresponding score.

2.3.3 Using Twitter in Economic

(Bollen et al. 2011), used Twitter as a source of knowledge for expecting changes in securities markets. The authors deal with 9,853,498 tweets gathered between 28 Feb and 19 Dec 2009. They used unsupervised learning approach to estimate the polarity of the opinions written in tweets. The authors used two lexicons OpinionFinder and GPOMS. OpinionFinder is a list of the subjectivity guides that is fragment of the OpinionFinder software, and GPOMS is an extension of the POMS lexicon created by the authors. The authors compare results generated with changes in Dow Jones Industrial Average (DJIA) stock markets. Finally, the authors propose that results obtained with OpinionFinder does not predict exactly the value of DJIA, while several GPOMS emotional are able to predict changes in the DJIA.

(Qaisi & Aljarah 2016), used sentiment analysis of the topmost cloud service suppliers specifically; Amazon and Microsoft Azure to examine their clients' feelings. Where they were relying on the data collected from twitter to sentiment analysis about certain products, brands and service. The end result presents that Microsoft Azure has 65% positive tweets compared 45% for Amazon.

(Shukri, Yaghi, Aljarah, & Alsawalqah 2015), applied sentiment analysis models to break down people groups' assessments and survey around three of most of the car industry companies to extract polarity and feelings of customers around each company. The outcomes presented that Audi was higher positive polarity (83%) than different organizations. Furthermore, the negative polarity of Audi was less than all other companies (BMW and Mercedes).

(Hodeghatta 2013), proposed an approach to analyze the tweets of Hollywood movies and understand the sentiments, emotions, and opinions expressed by the people across different parts of the world. The experimental performed on Twitter to classify messages of Hollywood movies as positive or negative and analyze sentiments through various areas of various countries. The outcomes showed that different regions express different sentiments depending on the nature of the movie and how the movies impact cultural sentiments.

2.3.4 Using Twitter in prediction of events

One the principal topics of importance in Twitter studies is the guess of events based on temporal chain. For example, (Asur & Huberman 2010) attempt to determine the usefulness of what is written on Twitter, as regards the prediction of coming events. The authors used the box office earning of a movie as a case study. To begin with, they endeavored to foresee the income of a film from its initially end of the week in the theaters from the proportion of tweets posted on the film in the week prior to its release. This drove them to formulation of the term tweet-rate, which is identified as the quantity of tweets every hour that indicate to a specific movie. The experiment indicated a solid association between the tweet-rate and incomes in the end of the first week. With the purpose of additional enhancing the result of the prediction, they included to their liner regression model the ratio of positive and negative tweets. To compute this ratio, they initially needed to decide the polarity of the tweets of each film by using a supervised learning approach. A fundamental basic for supervised learning is to have a set of training data. They applied the DynamicLMClassifier algorithm from the LingPipe NLP (Alias-i. 2008), to generate the classification model. The authors finished up with the popularization of their model for predicting commercial results and any product or service.

(Barnaghi, Ghaffari, & Breslin 2016) attempted to demonstrate the correlation among Twitter sentiment and events that have passed off at some point of the FIFA world cup 2014. The authors used a famous machine learning technique to achieve sentiment polarity for certain main actions based on Twitter data. The experimental results showed the positive and negative reaction of people towards some events and how it can change based on incidents during those events.

2.3.5 Using Twitter in Public health

(Paul & Dredze 2011), proposed work that compared Twitter posts with influenza ranks in the United States. The Ailment Topic Aspect Model find out mentions of many diseases such as allergies, obesity and insomnia after it had been applied to more than million and a half tweets related to public health. As well as the researchers combined prior knowledge into this model to perform a variety of tasks such as tracing illnesses over times and determining behavioural danger issues. The researchers were able to observe the presence of quantitative correlations with public health data and qualitative evaluations of model outcome.

Our approach is related to previous efforts as it also investigates political opinions towards Palestine. However, our study is distinguished in the following: First, it is the first study, to our knowledge, that explores the perception of Palestine by using data mining techniques over Twitter data. Second, we will explore the influence of features that have not been explored in previous studies and that are related to the Palestinian-Israeli conflict. For example, we will explore the impact of ethnicity on the attitude toward Palestine, the trade-off between the sympathy with Palestine and the support of Israel, and the influence of country opinion leaders. We may also explore the change in public opinion across time, and we will perform analysis at both individual and country levels.

Chapter 3

Methodology

Chapter 3
Methodology

3.1 Introduction:

In this chapter, the approach used to capture the public perceptions of Palestine from tweets is represented. The approach basically relies on the tweets' sentiment analysis in order to analyze and measure the perception towards Palestine. Sentiments are then categorised with respect to different features extracted from tweets such as the country of origin, leadership, time and ethnicity. The goal is to identify how the position towards Palestine has be affected by different features. The overall approach is depicted in Figure (3.1), and consists of three steps: data pre-processing, sentiment analysis, and feature extraction and analysis. The following subsections starts by describing the data collection process and the structure of the dataset. Then the approach is explained in detail. These steps are explained in detail in the following subsections.

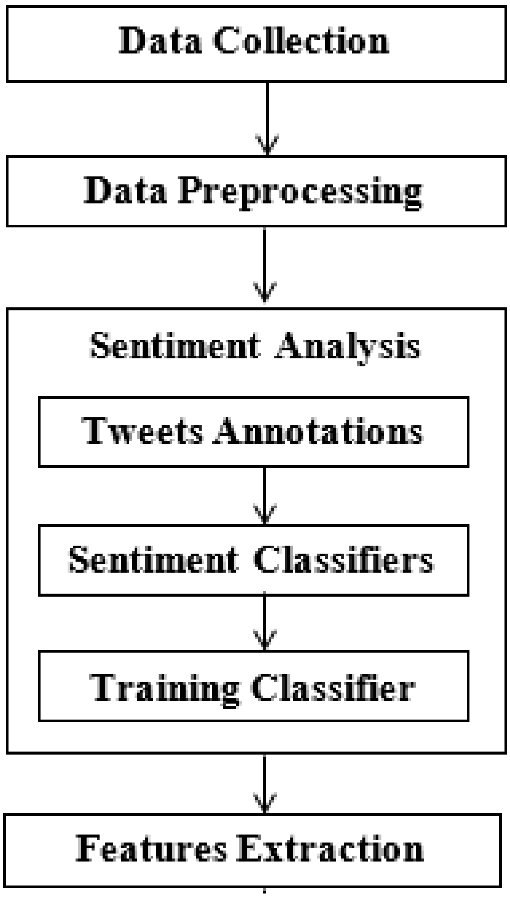

Figure (3.1): Methodology phases

3.2 Data Collection

Twitter API is a streaming API offered by Twitter for collecting tweets (Twitter 2017). Hypertext Transfer Protocol (HTTP) client libraries were also developed in different programming languages for consuming Twitter's streaming API. When we started the data collection process we used a Java library called twitter-4j (Yusuke 2007) to consume tweets. However, Twitter API has many restrictions that can result in insufficient data for our work. One of the most important restrictions is that it retrieves tweets for few days backward, usually seven days. In addition, a lot of features, such as the user name and country of origins, may be missing. Accordingly, we have used twitter search analytics and business intelligence tool called Followthehashtag (DNOiSE. 2017). Followthehashtag can help you search for tweets for a low price. A user can search for tweets posted through a specific period.

We performed a query-based search to collected tweets related to the Palestinian-Israeli conflict such as (Gaza, Gaza-strip, Jerusalem, Palestine-Israel conflict, Israel's occupation, Israel-Palestine, Hamas, Qassam, and Fatah). We collected 178,524 tweets posted by approximately 48,531 users during the period (Dec 20 2015 to Dec 31 2016) using Followthehashtag tool. Most of these tweets are from US, UK, Canada, Australia, Finland and some European countries. 89.78% of the collected tweets were in English.

For each tweet, the following information were retrieved: the tweet identifier, geographical location, and the date-time of submission, language, and textual content. Table (3.1) shows original column names.

Table (3.1): Structure of tweet

Column Name	Description
Date	Date of the tweet
User Name	@username
Nickname	Name shown in bio (“Real name”)
Tweet content	Text of the tweet
Favs	Favs amount
RTs	RTs amount
Latitude	Lat of apparent user location
Longitude	Long of apparent user location
Country	Name of the country
Place (as appears on Bio)	Declarated “place” in bio
Followers	Number of followers
Following	Amount of following
Tweet language	Language of tweet
Is a RT	If tweet is not original and is a retweet will be “True"
Hashtags	#Hashtags mentioned

The sample of tweets with the features we used can be found in Appendix A. The whole dataset can be found on the following link: https://github.com/odahroug2010/2017. Table (3.2) shows statistics about the collected tweets.

Table (3.2): Statistics about tweets

General information	**Total# of tweets**	178,524
	Number of users	48,530
	Duration	Dec 20 2015 to Dec 31 2016
	English tweets	89.78%
	Retweet	7948
	Avg. terms in each tweet	12.74
	Standard Dev. of terms	5.002
Location information	**No. of countries**	174
	Top sources of tweets	US, UK, Canada, Australia, Finland, and European
	Unknown recourse of tweets	28156
	Min. tweets by country	1
	Max. tweets by country	27490
	Avg. tweets by country	777.88024
	Standard Dev.	3363.68369

Note that 7948 tweets in total were retweets. Retweets were ignored so that only the original tweet is processed and analysed.

3.3 Tweet Pre-processing

Preprocessing is an essential part of analysis in order to preparing input data for manipulation. Tweets often have special characteristics that make the pre-processing of them different from that of ordinary text. Tweets are of limited length (140 characters at most), any they may contain special texts such as hashtags, URLs, emoticons, and usernames. For the pre-processing of tweets we used the same approach used in (Kechaou, Ammar, & Alimi 2011), and which is depicted in Figure (3.2).

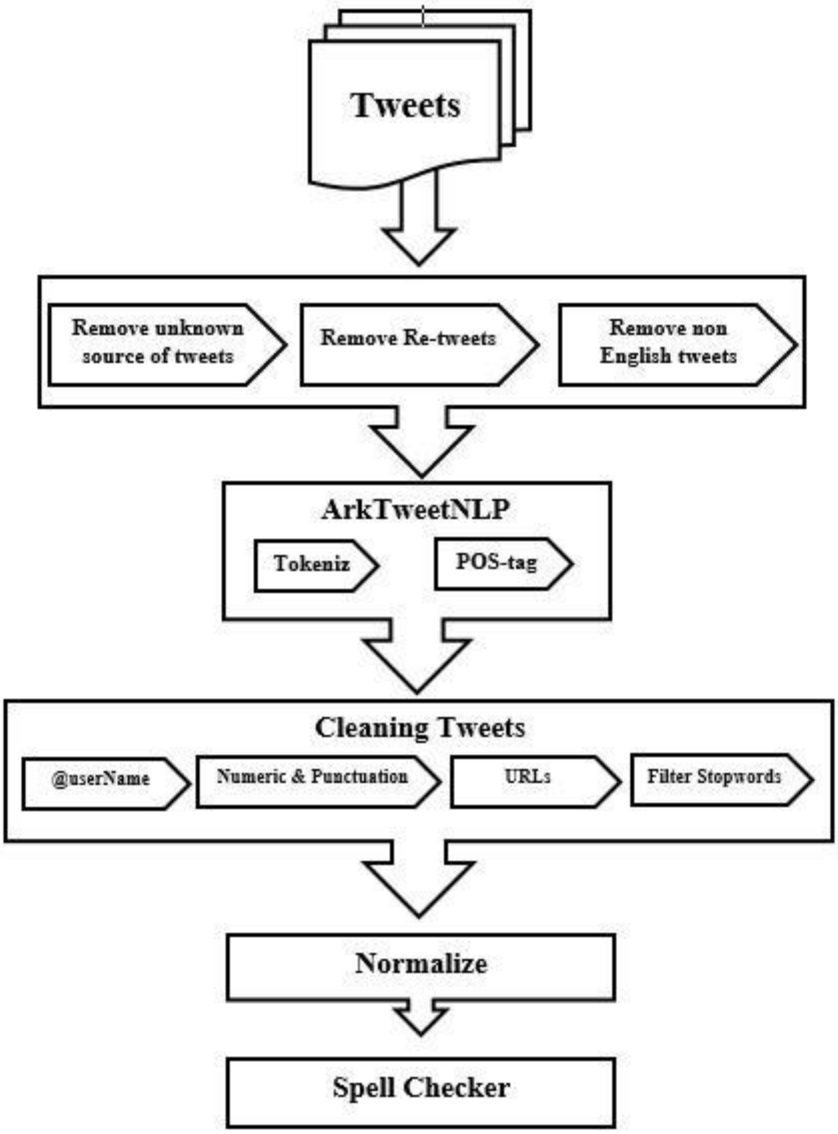

Figure (3.2): Pre-processing phases

3.3.1 Removing non-English Tweets

In our approach, we find 10.22% of tweets were written in non-English language and thus will be excluded.

3.3.2 Removing Unknown Source Tweets

Since our approach is concerned with knowing the western countries attitudes toward Palestine, the tweets of unknown sources, 28,156 tweets, are excluded.

3.3.3 Remove Re-tweets

In our approach, we find 7948 of tweets were retweeted, so we excluded from our dataset.

3.3.4 Tokenization and Tagging

Tokenization is the process that breaks text into words, phrases, symbols, or other significant parts called token for additional processing. Tagging or also called Part of Speech (POS) is the process of assigning a part of speech to each word and other token in sentence, such as nouns, verb, adjective. Tagging cannot be separated form tokenization over the pre-processing task.

Twitter allows users to write short messages, or tweets, which are cannot exceed 140 characters. For this reason, tweets had its own particular grammar and abbreviations in writing. Therefore, there is a need for specialized tokenizer that is able to recognize each token, the hashtag, the emoticons, and the URLs.

The following example illustrates the unique processing needed for tokenizing and tagging tweets. The tweet:

RT @Yaser I luv alwz CocaCola for 15 years :). #CocaCola http://c.u

Is tokenized and tagged as display in Table (3.3).

Table (3.3): POS annotations

Token	Annotated Tag	Description
RT	~	discourse marker, (indications of continuation of a message across multiple tweet)
@Yaser	@	at-mention (indicate another user as a reception of tweet)
I	O	pronoun (personal/WH; not possessive)
Luv	V	verb incl. copula, auxiliaries
Alwz	R	Adverb
CocaCola	^	proper noun
For	P	pre- or postposition, or subordinating conjunction
15	$	Numeral
Years	N	common noun

Token	Annotated Tag	Description
:)	E	Emoticon
.	,	Punctuation
#CocaCola	#	hashtag (indicates topic/category for tweet)
http://c.u	U	URL or email address

The previous example shows that traditional tokenizer and POS taggers may be inadequate for pre-processing tweets. Instead, there are several tokenizers that are developed specifically for informal and online conversational text including various non-standard lexical items and syntactic (Gimpel, Schneider, & O'Connor 2013). In our work, we used a library called ArkTweetNLP to tokenize and tag tweets in Java (Owoputi, O'Connor, Dyer, Gimpel, Schneider, & Smith 2013).

3.3.5 Cleaning Tweets

Twitter users prefer to a use set of symbols, abbreviations, and non-standard language in their tweets. A large number of these tokens should be excluded before further processing, to avoid an incorrect and misleading result when applying the sentiment analyser. In our approach, tweets were cleaned by removing the following parts:

- **Usernames:** usernames are references to user accounts within the tweet text. They are often preceded by @ sign.
- **Numeric expression:** all numerical expressions annotated with ($) tag in tweets are deleted.
- **Punctuation**s.
- **URLs:** all links (URLs) posted on tweets and annotated with (U) tag, which we have mentioned in Table (3.3), are deleted.
- **Filtering Stop-Words:** Stop-words generally indicate the most popular words in a language which have very little meaning, such as "and", "the", "a", "an", and similar words. They can be removed from a text since they do not affect the final sentiment of the text. The list of deleted stop words can be found in Appendix H.

3.3.6 Normalization

There are many linguistics' noise and abbreviations in text as a result of the ways in which texts are written. Normalization is the method of converting text into an original form as following steps:

- **Emoticons:** Emoticon is a representation of face appearance using punctuation marks, numbers and letters, usually written to express sentiment and opinions in text. Emoticons are important for sentiment analysis; thus, their meanings should be preserved and should not be removed from tweets. In our approach, we used a special dictionary that contains the most used emotions and its meaning in English (Gimpel et al. 2013), Table (3.4) illustrated some of these emotions. This table is used during the pre-processing step to match and replace each emoticon with its relevant meaning.

Table (3.4): Emotions

Emoticon	Word	Emoticons	Word
:)	Happy	(:	Sad
;)	Wink	:'(	Crying
>:o	Surprise	-_-	Sleeping
@_@	Amazed	>:(	Evil
o.O	Surprise	;D	Wink
:/	Annoyed	:))	Happy
-:(	Sad	:P	Cheeky
8D	Laughing	>.<	Annoyed
XD	Laughing	=|	Angry

- **Lowercase Letters:** all tweet letters converted to lowercase in order to improve string matching.

3.3.7 Spell Checker

Often, we note that some tweets contain incorrect words or an error in writing, and this will affect the result of sentiment analysis if these words remains as they. This step manipulates these words in the spell checker and substitute with the best match. To do this, we have used Jazzy Spell Checker for spell checking (IDZELIS 2005).

Table (3.5) illustrates some of tweets with spelling mistake before correction and after invocation method for spelling correction.

Table (3.5): Jazzy spell checker results

Before Correction	After Correction
I looooooove palestin	I love Palestine
i'm happi to vist univesity	i'm happy to visit university
i hopeee to vist jeruslem	i hope to visit Jerusalem

3.4 Sentiment Analysis

Sentiment analysis is the core step to identify attitudes toward Palestine. When sentiments are identified, tweets can be categorized based on different features. Therefore, results of subsequent steps will depend on the quality of sentiment analysis of tweets.

We assume that the tweet is an opinion, and therefore we need to know its polarity classification, which is positive, negative, or neutral. To achieve this, we used a supervised approach for sentiment analysis which is explain in what follows.

It is important to notice that the aim of sentiment analysis in this work is to identify the political stance towards Palestine and not the user feeling. Therefore, it is important to differentiate between the emotion and the political sentiment, which may be different for the same tweet. A tweet may reflect a negative emotion (e.g. sadness) but it may carry positive attitude towards Palestine. For example, the sentence: *"I feel sorry for the Palestinian children who were arrested and are now in Israeli prisons"* should have a positive polarity in our case because it shows support for Palestine, despite the fact that it carries a sad sentiment. Similarly, tweets that evoke a positive emotion towards "Israel" (e.g. "*I love Israel*") should have a negative polarity from the perspective of Palestine.

3.4.1 Choosing a Sentiment Classifier

Initially, we decided to start with one of the off-the-shelf and pre-trained sentiment analysers and use it to measure the polarity of tweets. We started with three sentiment analysers that are:

- Stanford CoreNLP (Manning et al. 2014).
- SentiStrength (Thelwall, Buckley, & Paltoglou 2012).
- LingPipe (Alias-i. 2008).

Our purpose was to test these analysers and choose the optimal sentiment analyser that can give the best results from our data set.

To evaluate and compare the performance of the above sentiment analysers, we should have a labelled dataset to be used as a gold standard. Since our collected tweets does not include predefined sentiments, we decided to pick a number of tweets and label them manually with relevant polarity (positive, negative, or neutral). These labelled tweets will be then used to evaluate the sentiment analysers. 1000 tweets from our dataset were chosen and were given to two human subjects to label them separately.

In general, the labelling of tweets was done as the following:

- Tweets that include appreciation, praise, glorification or support for Palestine or the Palestinian issue were label as positive. For example, idioms like *"Free Palestine"*, or *"It is called Palestine not Israel"* have positive polarity.

- Tweets that show solidarity and sympathy with Palestine or Palestinian victims were labelled as positive. For example, idioms like *"Please donate for the Children of Gaza"* or *"Save Palestinian Children from Israeli Genocide"* should be labelled as positive.

- Tweets that contain idioms denoting negative attitude towards what so call *"Israel"* are considered positive from the perspective of the pro-Palestinian point of view. For example, idioms like *"Israeli Occupation"*, *"Zionist massacres"* or *"Apartheid wall"* all carry positive sentiment towards Palestine, and thus were labelled as positive.

- Tweets that show clear support for or sympathy with what so called *"Israel"* or any of its deeds were labelled as negative. For example, idioms like *"I love Israel"*, or *"Israel has the right to defend itself"* all carry positive attitude towards *"Israel"* and negative attitude towards Palestine, and thus were labelled as negative.

- Tweets that use Israeli naming conventions, such as *"Judea and Samaria"* and *"IDF army"*, *"Palestinian terrorists"* were treated as negative sentiments since they adopt a pro-Israel stance.

After analysing labels received from the two subjects and ignoring disagreements, we ended up with 882 tweets, of which 426 were positive, 319 were negative, and 137 were neutral. Table (3.6) shows statistics about the labelled tweets.

Table (3.6): Statistics about the tweets labelled by human subjects

Sentiment	Number of sample
Positive	426
Negative	319
Neutral	137
Total	882

3.4.2 Performance Metrics

In this section, we will describe the most commonly performance metrics and that will be used in our study.

3.4.2.1 Accuracy

Accuracy is estimated by dividing the total correctly classified positives and negatives by the total number of samples (Olson & Delen 2008), Equation (3.1).

$$Accuracy = \frac{N(correct\ classifications)}{N(total\ number\ of\ sample)} \tag{3.1}$$

3.4.2.2 Precision, Recall, and F-measure

Precision and recall are measurements for relevance usually used in pattern recognition and information retrieval with binary classification (Olson et al. 2008). Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved, while recall is the ratio of the number of relevant records retrieved to the total number of relevant records.

High precision means that an instances had been retrieved are finally more relevant results than irrelevant, while high recall means that the most of relevant results are retrieved (Olson et al. 2008).

Precision (P) is defined as the number of true positives T_P over the number of true positives plus the number of false positivesF_P, Equation (3.2).

$$P = \frac{T_P}{T_P + F_P}$$

(3.2)

Recall (R) is defined as the number of true positive T_P over the number of true positives plus the number of false negativeF_n, Equation (3.3).

$$R = \frac{T_P}{T_P + F_n}$$

(3.3)

F-measure is a measure of a test's accuracy for binary classification. The F-measure can interpreted as a weighted average of the precision and recall, Equation (3.4), where an F-measure score reaches its best value at 1 and worst at 0 (Olson et al. 2008).

$$F - measure = \frac{2 * P * R}{P + R}$$

(3.4)

3.4.3 Evaluation of Pre-Trained Sentiment Classifiers

As mentioned earlier, we picked three pre-trained sentiment classifiers (Stanford CoreNLP, SentiStrength, and LingPipe), and aimed to choose the classifier that gives the best results from our labelled tweets. We evaluated the classifiers by comparing the performance metrics. The results of this comparison are presented in Table (3.7).

Table (3.7): Classifiers performance

S. No	Classifier	Performance (Accuracy)	Precision	Recall	F-measure
1	Stanford CoreNLP	8.1%	30.6%	22.6%	26.1%
2	SentiStrength	7.9%	42.2%	27.8%	33.5%
3	LingPipe	31.2%	35.6%	30.5%	32.9%

Table (3.7) shows that the performance metrics values for the three sentiment classifiers. It is obvious that all classifiers performed poorly and did not generate satisfactory results that can build upon. These poor results can be explained by the fact that these classifiers are trained to identify the mental feelings or emotions rather than the political sentiment. As explained in Section 3.4.1, judging a sentiment to be positive or negative depends on the side you stand by. Therefore, the same idiom may be translated differently based on the context. For example, the idiom: "I feel sorry for the Palestinian people☹" conveys a feeling of sadness and sympathy, and was classified as "Negative" by the above classifiers, despite that it carries a positive attitude towards the Palestinian case. Table (3.8) illustrates the differences between the manual labelling and classification results for sample tweets.

Table (3.8): Sentiment results

Tweet	Manual labeled	Classifier	Result
I extremely love Israel	Negative	Stanford CoreNLP	Positive
		SentiStrength	Positive
		LingPipe	Positive
It is called Palestine, NOT Israel	Positive	Stanford CoreNLP	Negative
		SentiStrength	Neutral
		LingPipe	Neutral

Based on the above results, it was decided that pre-trained sentiment classifiers are inadequate for political sentiment analysis. Instead, we decided to train a sentiment classifier on a manually-labelled dataset.

3.4.4 Training the classifier

Of the three-former sentiment classifiers we listed in Section 3.4.1, a LingPipe classifier was used in our experiment because it can be easily configured and trained with custom datasets. It classifies texts by using a language model on character sequences, and the execution uses the 8-gram language model.

The 882 tweets, which were labelled manually as explained in Section 3.4.1, were randomly split into two parts: 80% of tweets were used for training, and 20% were used for testing. 10-fold cross validation was performed (This is done internally by Apache Spark). Table (3.9) shows the accuracy of the LingPipe sentiment analyzer is (81.21%). Based on this result, the trained LingPipe sentiment analyzer will be used to measure the polarity of tweets of our dataset.

Table (3.9): LingPipe accuracy

Experiment	Data set	No. of tweets				Accuracy
		Positive	Negative	Neutral	Total	
Experiment	Training	356	249	112	717	**81.21%**
	Testing	70	70	25	165	
	Correct	**64**	**60**	**10**	**134**	

In addition, we measured the precision and recall for each class by creating confusion matrix as shown in Table (3.10). The matrix shows that the analyser achieves good results with positive and negative tweets, but the performance was low with neutral tweets. However, we think that the low performance in the case of neutral tweets will not have high impact on results due to the low number of neutral tweets in general.

Table (3.10): Precision and recall

	Label Positive	Label Negative	Label Neutral	Total Predicted	Precision	Recall
Predict_Positive	**64**	4	2	**70**	**91.4%**	**80.0%**
Predict_Negative	6	**60**	4	**70**	**85.7%**	**86.9%**
Predict_Neutral	10	5	**10**	**25**	**40.0%**	**62.5%**
Total Label Class	**80**	**69**	**16**		**72.4%**	**76.5%**

3.5 Conclusion

This chapter described the approach we used to collect, clean and pre-process tweets related to Palestine and the Palestinian-Israel conflict. It also presented the method used to perform sentiment analysis of the tweets to identify the political attitude towards Palestine. The validity of the sentiment classifier to measure sentiments towards Palestine was also assessed and discussed.

Chapter 4

Data Analysis and Results

Chapter 4
Data Analysis and Results

4.1 Introduction

In the previous chapter, the political sentiments towards Palestine were measured from Twitter data. In this chapter, the data analysis will be performed in order to generate new insights into the perception of western media towards the Palestinian issue.

Data analysis was performed at two levels based on the approach shown in Figure (4.1): country-level analysis and individual-level analysis.

The purpose of country-level analysis is to explore the country's overall interest in and attitude towards Palestine by:

- Identifying counties that generate the most Palestine-focused tweets by aggregating tweets at the county level. The aim is to determine counties that tweet most about Palestine regardless of the friendliness levels.
- Measuring the friendliness of each country towards Palestine. Friendliness can be determined from the polarities of tweets of each country.
- Analysing time series data to investigate how the country's attitude towards Palestine changes over time.

The individual-level analysis aims to analyze data based on the activity of individuals. We will perform the following analysis:

- Capture the attitudes of opinion leaders, and measure their friendliness towards Palestine. The term "Opinion leaders" refers to users on social media who have a large number of followers and often post frequently (Khan, Ata, & Rajput 2015).
- Analyse the influence of ethnicity on the public opinion, and the potential relation between the ethnicity of users and their attitudes towards Palestine.

The above analysis will be performed based on the sentiments measured from the collected tweets.

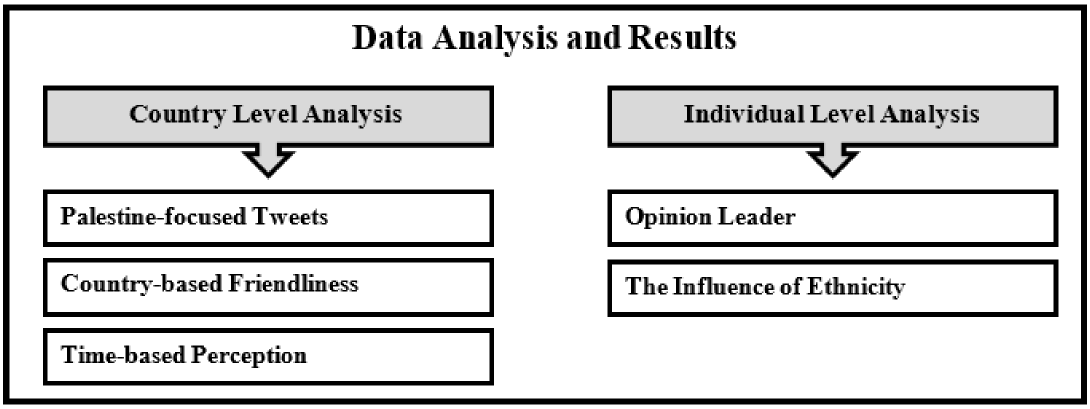

Figure (4.1): Data analysis level

4.2 Feature Extraction

In our study, several features need to be extracted from the collected tweets to perform both country-based and individual-based analysis. These features are listed below. The extraction and measurement of these features will be explained in subsequent section.

- **Polarity:** polarity is the sentient score of the tweet, which determines the classification of the tweet (e.g. positive, negative, or neutral). Polarities of tweets have been measured by using the sentiment analyser as detailed in Chapter 3. All other features will be derived from the polarities of tweets.

- **Friendliness:** The friendliness of a country is measured by calculating the average polarity of tweets posted by users in the country. Similarly, friendliness of an individual is the average polarity of tweets posted by the user. To compute the average polarity, we interpreted the three sentiment values: positive, neutral, negative into +1, -1, 0 respectively. Then, the friendliness for a country Fc is computed using the following equation (4.1):

$$\boldsymbol{Fc} = \frac{\sum \boldsymbol{Polarity}(\boldsymbol{t_i})}{\boldsymbol{n}} \times \mathbf{100}$$

(4.1)

Where n is the number of tweets attributed to the country c. t_i is a tweet posted in the country c.

- **Leadership:** this feature is used to identify opinion leaders. In this study, Twitter users who have the most number of followers are treated as opinion leaders.

Measurement of this feature and its use for data analysis will be explained in Section 4.4.1.

- **Ethnicity:** In this study, the ethnicity of the Twitter user will be identified from the user's name or nickname. The detailed extraction of this feature and its potential for data analysis will be explained in Section 4.4.2.

4.3 Country-Level Analysis

This section starts by classifying countries based on the volumes of Palestine-focused tweets. The total number of tweets related to Palestine from each country indicates the level of concern and awareness in that country about the Palestine-Israel issue. Afterwards, country-based friendliness scores are calculated, and countries that are most friendly with Palestine are identified. Finally, time-based analysis is performed to explore changes in attitude over time.

4.3.1 Palestine-focused Tweets

The volume of tweets that can be attributed to each country should be identified. Each tweet often comes with geo-information that can help identify its country of origin. One attribute is called "country" and it should be set with the country code. For example, tweets posted from the UK have the country value "GB". Another attribute is named "place" and is often set with the state or city name. However, the attributes "country" and "place" may be missing for many tweets, and they can be identified only if they are set in the user profile. Tweets can also have geocoding attributes named "Latitude" and "Longitude". These attributes are set to valid values for tweets posted from devices with enabled GPS service.

Only tweets that have either a valid country name or valid latitude and longitude values are used. In total, 150,368 tweets are geocoded with either a country name or latitude –longitude values.

To recognize the country name, the "country" and "place attribute" are first checked. If they are empty, we refer to latitude and longitude values, and map them to the country name by using the Google Map[(1)] geocoding service. The Google Map geocoding services takes latitude and longitude values as input and returns the county

---

(1) http://maps.google.com

name as output. After assigning tweets to countries, countries that that have a number of tweets less than 0.1% of total number of tweets were neglected.

Table (4.1) shows the top twenty countries in terms of the number of Palestine-focused tweets. Canada, the UK and the US generated most Palestine-focused tweets. The bottom countries were Slovenia, New Zealand, and Austria. Figure (4.2) illustrates the results on a geographical map. The complete results can be found in Appendix B.

Table (4.1): Palestine-focused tweets per country

No.	Country	Country code	Focused Tweets
1	Canada	CA	27,490
2	United Kingdom	GB	23,010
3	United States	US	20,125
4	Jersey(1)	JE	11,739
5	Ecuador	EC	9,342
6	Finland	FI	3,654
7	Australia	AU	3,125
8	Netherlands	NL	2,646
9	India	IN	1,445
10	France	FR	1,215
11	Ireland	IE	971
12	Pakistan	PK	971
13	Germany	DE	830
14	Greece	GR	820
15	South Africa	ZA	717
16	Japan	JP	642
17	Denmark	DK	639
18	China	CN	636
19	Italy	IT	577
20	Indonesia	ID	521

(1) https://en.wikipedia.org/wiki/Jersey

Figure (4.2): Palestine-focused tweets per country

When considering the number of population, Jersey, Canada and Finland generated the most tweets per capita. The bottom countries were Slovenia, New Zealand and Austria. Table (4.2) shows the top 20 countries in terms of Palestine-focused tweets per capita. Figure (4.3) illustrates the results on a geographical map. The complete results can be found in Appendix C.

Table (4.2): Palestine-focused tweets per capita

Country	Code	Focused tweets by country	Population (2016)	Focused tweets by capita
Jersey	JE	11,739	164,541	7.134392036
Canada	CA	27,490	36,289,822	0.075751267
Finland	FI	3,654	5,503,132	0.066398553
Ecuador	EC	9,342	16,385,068	0.057015326
United Kingdom	GB	23,010	65,788,574	0.034975678
Ireland	IE	971	4,726,078	0.020545577
Netherlands	NL	2,646	16,987,330	0.015576315
Australia	AU	3,125	24,125,848	0.012952913
Denmark	DK	639	5,711,870	0.011187229
Slovenia	SI	179	2,077,862	0.008614624
Greece	GR	820	11,183,716	0.007332089
United States	US	20,125	322,179,605	0.006246516
Switzerland	CH	381	8,401,739	0.004534775
Belgium	BE	399	11,358,379	0.003512825
Sweden	SE	293	9,837,533	0.002978389
Portugal	PT	249	10,371,627	0.002400781
Austria	AT	172	8,712,137	0.001974257
Kazakhstan	KZ	352	17,987,736	0.001956889
France	FR	1,215	64,720,690	0.001877298
South Africa	ZA	717	56,015,473	0.001280003

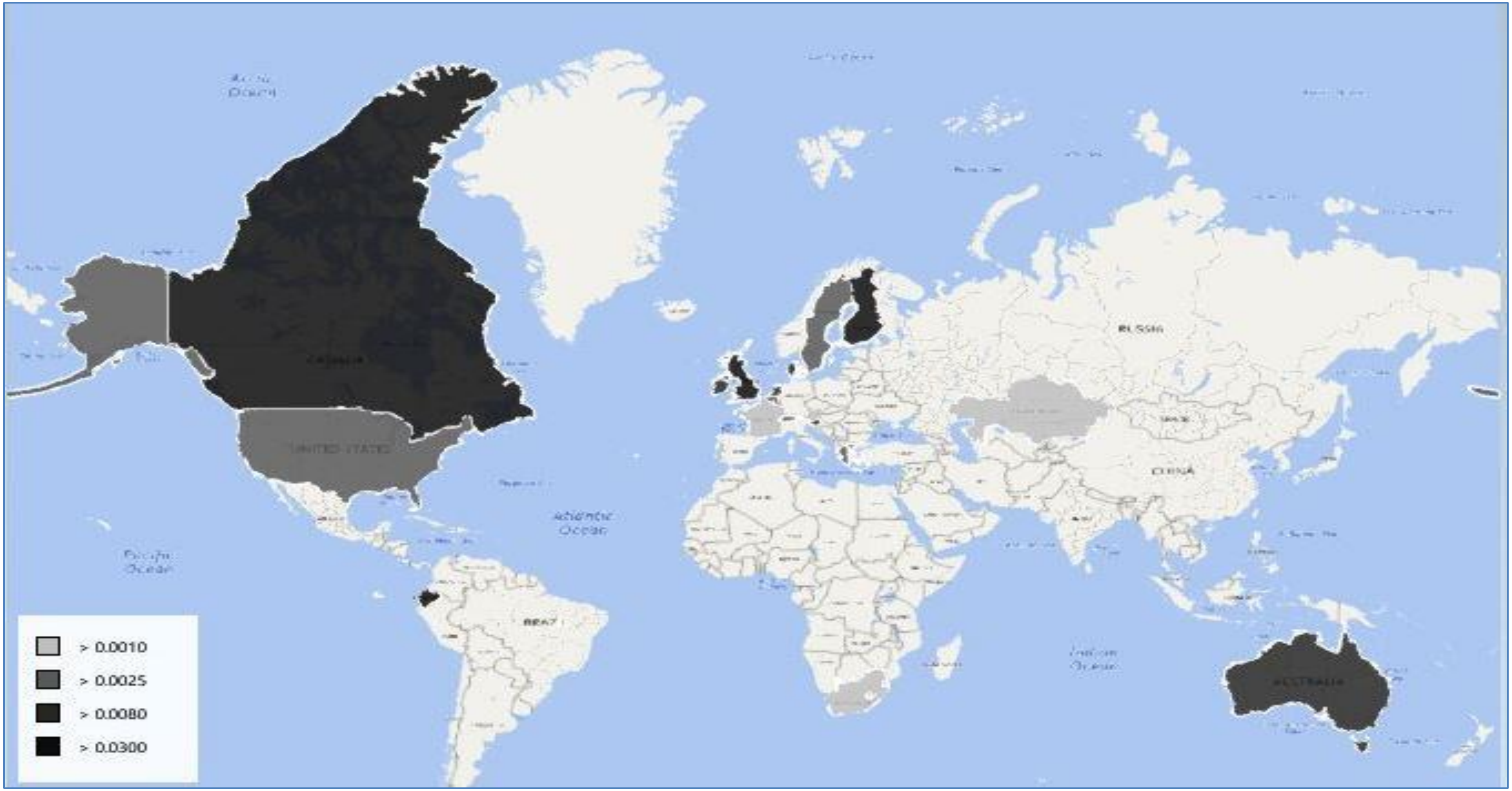

Figure (4.3): Palestine-focused tweets per capita

To get insight into the validity of Palestine focused tweets per country, we tried to compare our result with country index generated from Google Trends[1] . The Google index will indicate the frequency at which people in the country searched for the term during the specified period of time. We used Google index to measure the frequency at which the people in each country, top 38 Palestine-focused tweets, for the term "Palestine" from January 2016 to December 2016. To make easy comparison, we log-transform count of tweets tweeted per country to be comparable with values from Google index. The results can be found in Appendix D. We plot the Google index as the x axis and plot our Twitter index as the y axis. The result is reported in Figure (4.4).

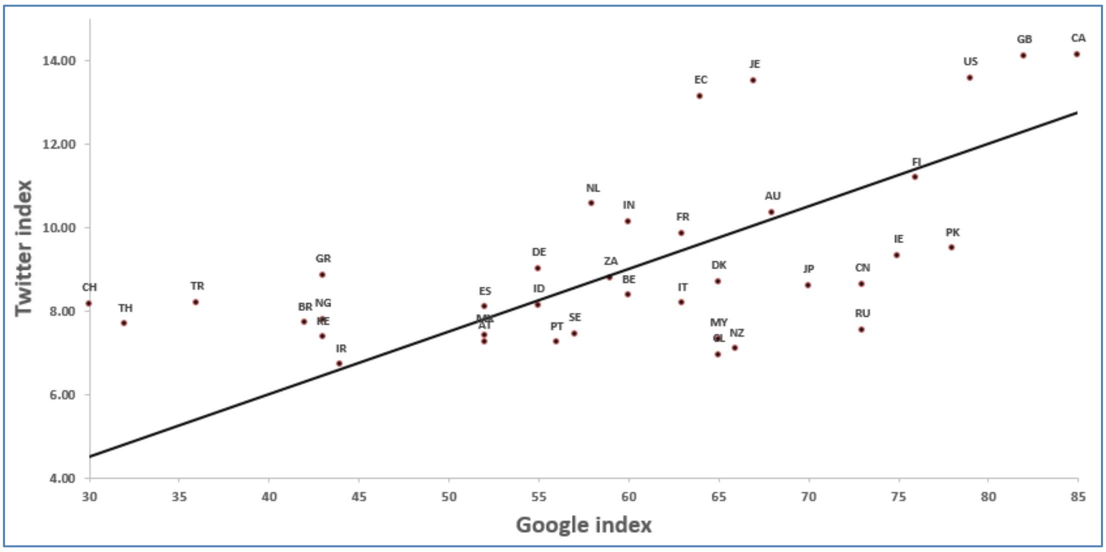

Figure (4.4): Compare Twitter index with Google index

We then measured the coefficient of correlation between the Google index value and Twitter index. The result is 0.5435, which is considered high correlation factor between them.

[1] Google Trends, https://trends.google.com/trends/?hl=en

4.3.2 Country-based Friendliness

The country-based friendliness is calculated by using Equation (4.1). It is the average sentiment scores for each country multiplied by 100. Note that the friendliness score for each country can be positive, negative, or zero denoting a positive, a negative, or a neutral attitude respectively.

Table (4.3) shows results for the top twenty countries in terms of friendliness. Figure (4.5) shows the friendliness scores for top twenty countries, while Figure (4.6) shows the standard deviation values for top twenty countries. Figure (4.7) showcases the friendliness scores on a world map. The complete results can be found in Appendix E.

Table (4.3): Top twenty country with respect to friendliness

No.	Country	Focused Tweets	Positive	Negative	Neutral	Friendliness	St. Dev
1	Finland	3,654	3,177	401	76	**75.97**	**0.63**
2	Brazil	382	184	118	80	**17.28**	**0.87**
3	Thailand	262	127	89	46	**14.50**	**0.90**
4	Japan	642	308	272	62	**5.61**	**0.95**
5	Netherlands	2,646	1,182	1,081	383	**3.82**	**0.92**
6	France	1,215	440	457	318	**-1.40**	**0.86**
7	Greece	820	317	338	165	**-2.56**	**0.89**
8	Nigeria	315	104	118	93	**-4.44**	**0.84**
9	Italy	577	207	235	135	**-4.85**	**0.87**
10	Islamic Republic of Iran	218	80	96	42	**-7.34**	**0.90**
11	Portugal	249	91	112	46	**-8.43**	**0.90**
12	China	636	191	249	196	**-9.12**	**0.83**
13	Chile	206	77	96	33	**-9.22**	**0.91**
14	United Kingdom	23,010	8,281	11,344	3,385	**-13.31**	**0.91**
15	Ecuador	9,342	372	1,762	7,208	**-14.88**	**0.45**
16	Indonesia	521	149	234	138	**-16.31**	**0.84**
17	Sweden	293	90	139	64	**-16.72**	**0.87**
18	Turkey	383	88	157	138	**-18.02**	**0.78**
19	Kenya	216	58	97	61	**-18.06**	**0.83**
20	India	1,445	296	576	573	**-19.38**	**0.75**

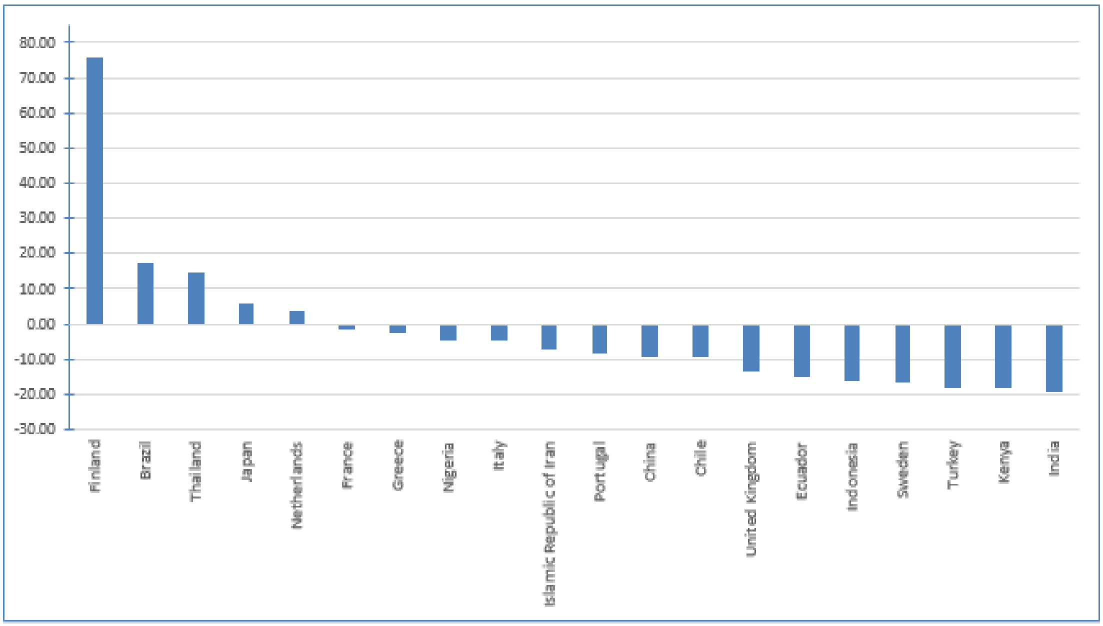

Figure (4.5): Friendliness country

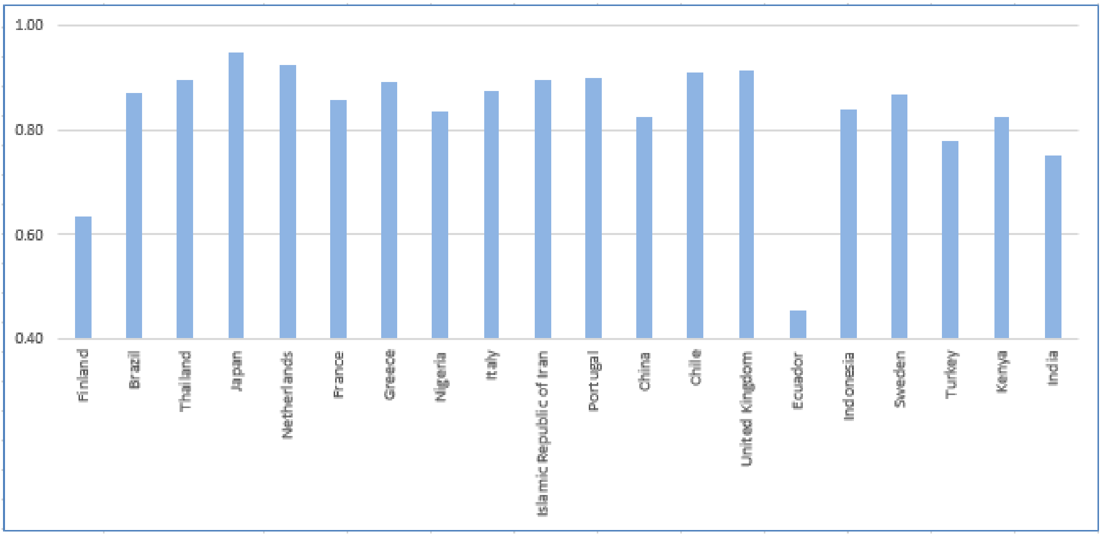

Figure (4.6): Standard deviation values

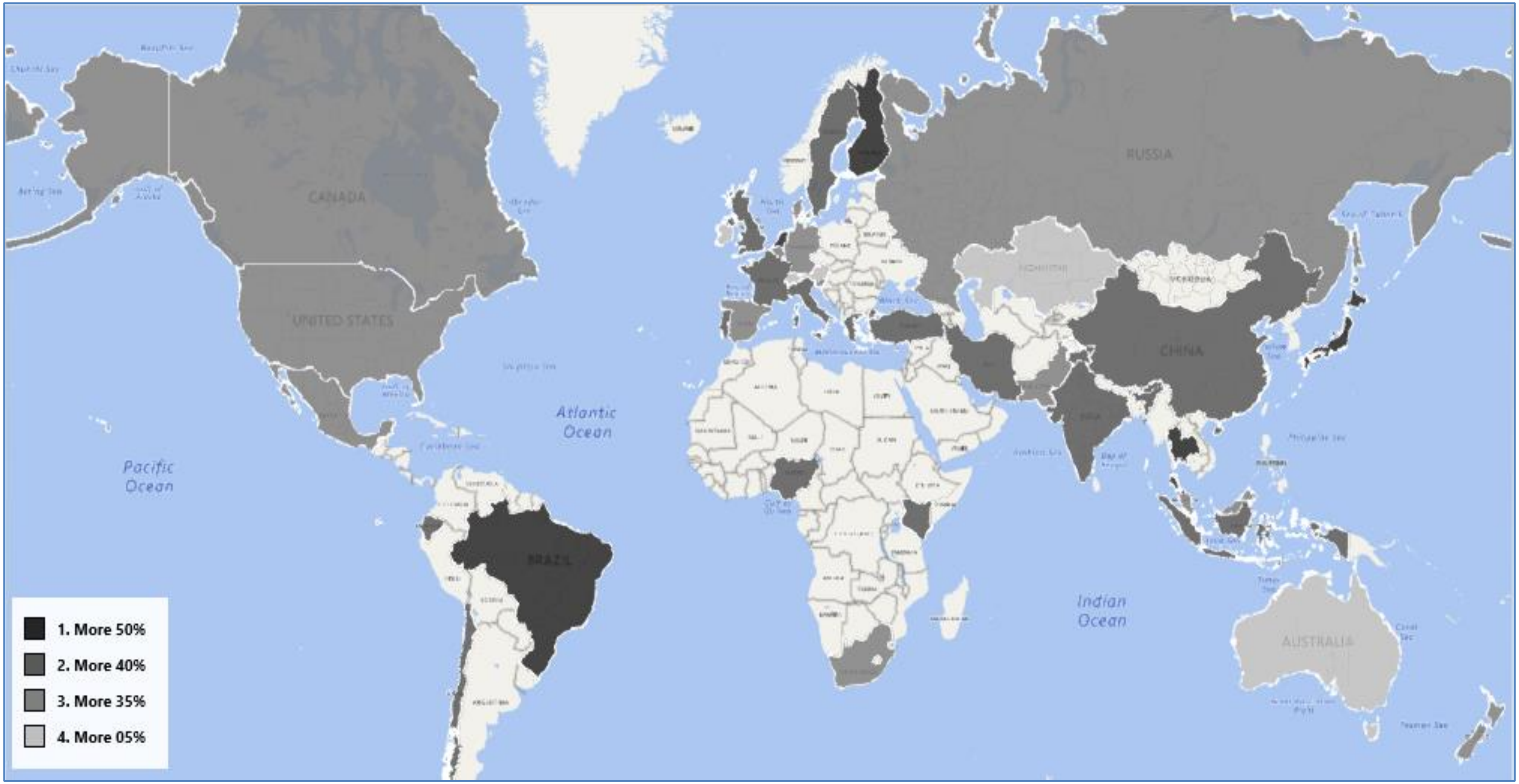

Figure (4.7): Friendliness scores on map

The friendliest countries were Finland, Brazil, and Thailand. The least friendly countries were Switzerland, Austria, and Kazakhstan. Of the top twenty countries, only five countries had friendliness scores over zero, while the rest got below-zero scores. This result indicates that the public opinion is still highly negative towards Palestine even in the top friendly countries. Several countries like France, Greece, Nigeria, and Italy got close to zero friendliness scores.

Referring to the distribution of sentiments and the standard deviation in Table (4.3), a high divergence of attitudes can be observed for most countries. For countries like France, Italy and the UK, the numbers of positive and negative tweets were mostly comparable, while the number of neutral tweets were smaller. The small number of neutral voices in these countries indicates the large polarization in public opinion towards the Palestinian issue.

Table (4.4) shows the friendliness scores for the top twenty countries in terms of Palestine-focused tweets. Countries in North America like the US and Canada have low friendliness scores despite the large number of Palestine-focused tweets. This result could be reasonable taking into consideration the strong influence of Jewish lobbies in these countries as compared to the influence of Jewish lobbies in European countries.

Table (4.4): Friendliness scores for the top twenty countries in terms of Palestine

No.	Country	Focused Tweets	Positive	Negative	Neutral	Friend liness	St. Dev
1	Canada	27,490	7,448	14,164	5,878	**-24.43**	0.85
2	United Kingdom	23,010	8,281	11,344	3,385	**-13.31**	0.91
3	United States	20,125	4,762	10,203	5,160	**-27.04**	0.82
4	Jersey	11,739	3,428	6,851	1,460	**-29.16**	0.89
5	Ecuador	9,342	372	1,762	7,208	**-14.88**	0.45
6	Finland	3,654	3,177	401	76	**75.97**	0.63
7	Australia	3,125	754	1,862	508	**-35.46**	0.84
8	Netherlands	2,646	1,182	1,081	383	**3.82**	0.92
9	India	1,445	296	576	573	**-19.38**	0.75
10	France	1,215	440	457	318	**-1.40**	0.86
11	Pakistan	971	251	483	237	**-23.89**	0.84
12	Ireland	971	230	593	148	**-37.38**	0.84
13	Germany	830	225	416	189	**-23.01**	0.85
14	Greece	820	317	338	165	**-2.56**	0.89
15	South Africa	717	171	370	176	**-27.75**	0.82
16	Japan	642	308	272	62	**5.61**	0.95
17	Denmark	639	177	322	140	**-22.69**	0.85
18	China	636	191	249	196	**-9.12**	0.83

No.	Country	Focused Tweets	Positive	Negative	Neutral	Friend liness	St. Dev
19	Italy	577	207	235	135	**-4.85**	0.87
20	Indonesia	521	149	234	138	**-16.31**	0.84

4.3.3 Time based Perceptions

One of the motivations of this study is to explore how the attitudes towards Palestine varies over time. As illustrated in section 3.2, each tweet in our dataset is associated with a timestamp that specifies when the tweet was posted. Therefore, tweets can be treated as time series that can be analyzed to extract meaningful patterns.

Due to the variations among countries, utilizing the whole volume of tweets for time series analysis can result in a large variance. Therefore, the time series analysis was carried out only for the top three countries in terms of the number of posted Palestine-related tweets. These countries are United Kingdom (GB), United State (US), and Canada (CA).

Figure (4.8) shows how the friendliness scores of these countries have changed over a year: from January 2016 to December 2016. It is obvious that the attitudes of the three countries fluctuate up and down, and they almost have the same pattern. Friendliness scores were low in the first half of the year, before rising up to a peak value in June-July. Attitudes then went down again, then went up at the end of September, before going down again.

To understand these results, we tried to link the time-based changes with the significant events that took places over the year and that are related to Palestine. These events are easily recognizable by searching the news archives on the Web. Table (4.5) shows statistics about the main events that may have significantly influenced the attitude towards Palestine during 2016. For each event, the number of related tweets and some examples are also illustrated in Table (4.5).

The declining attitude in the first quarter of 2016 may be explained by the stabbing spree that took place in Jerusalem and other Palestinian cities. A total number of 354 tweets related to the stabbing incidents were tweeted in the three countries during the first quarter of 2016. The rising attitudes towards Palestine in June-July 2016 may be attributed to the Israeli demolitions that took place in July 2016 and resulted in the displacement of dozens of Palestinians[(1)]. In addition, the press releases that accused "Israel" of forcing Palestinians

[(1)] http://www.aljazeera.com/news/2016/07/israeli-demolitions-displace-dozens-palestinians 160713124336539.html to withstand cruel and inhuman conditions at its borders have also grabbed attention during June 2016[(1)(2)].

Another rise of attitude towards Palestine was observed in September 2016, and can be explained by the reaction over the death of Shimon Peres, the former Israeli Prime minister who is widely considered as a war criminal by pro-Palestinians[(3)]. In total, 571 tweets referring to "Shimon Perez" were tweeted from these countries during September-October 2016, most of which had positive polarity. In addition, the UNESCO resolution on 12th October 2016 that condemned the Israeli policies around Al-Aqsa Mosque compound also got considerable attention in social media[(4)(5)]. 332 tweets related to the UNESCO resolution were tweeted from the three countries during October-November 2016.

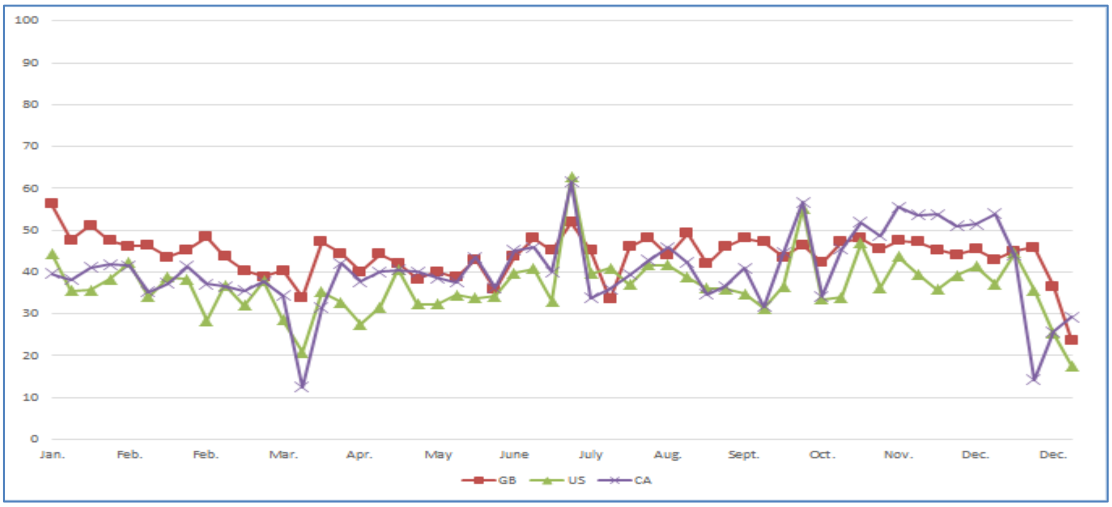

Figure (4.8): Time-based perceptions

(1) http://www.independent.co.uk/news/world/middle-east/israel-border-crossing-checkpoint-palestinians-west-bank-btselem-a7106486.html
(2) http://www.btselem.org/press_releases/20160727_house_demolitions_in_area_c
(3) http://www.independent.co.uk/news/world/middle-east/jerusalem-mayor-palestinians-animals-terror-attack-two-killed-meir-turgeman-a7355116.html
(4) https://www.middleeastmonitor.com/20161013-unesco-vote-no-link-between-al-aqsa-and-judaism/
(5) http://www.aljazeera.com/news/2016/10/palestinians-unesco-vote-al-aqsa-compound-161015133808135.html

Table (4.5): Statistics about Palestine-related events in 2016

Event	No. of tweets referring to the event	Sample tweets
Stabbing Spree	354	• The Israeli occupation forces shot injured a Palestinian young man after alleged stabbing attempt in Jerusalem • #Israel's abuse of Palestinians can NEVER justify stabbing 13-year-old girl Palestine. Stop killing civilians. • @StanleyCohenLaw if Israel wants to end the stabbings end the occupation. OR arm the Palestinians and make it a fair fight
Demolitions	153	• Israeli forces destroy five Palestinian homes leaving 27 Palestinians homeless. • Israel to destroy Palestinian school in West Bank, UN demands halt to demolitions. • Wibisono at UNHRC calls on Israel to halt settlement expansion and stop Palestinian home demolitions.
Cruel and inhuman condition	346	• "Oppression, Occupation, expansion and the cruel inhuman treatment of Palestinians; Jews/blacks by #Israel remains in place but not for long". • Israel's Blockade of Gaza Is Killing Women with Breast Cancer. #BDS to end this cruel oppressive occupation. • Sanders offered to work to end Israel's inhumane occupation of Palestine. We must demand Hillary Clinton do the same
Shimon Peres	571	• Obama visit to Shimon Peres funeral most likely to legitimise the war crimes. • Abbas and Netanyahu shake hands at Shimon Peres funeral.
UNESCO	332	• Opinion: UNESCO must go one step further with a cultural boycott of Israel. • Jerusalem under Occupation: Israel outraged by UNESCO resolution on Jerusalem sites.

4.4 Individual-Level Analysis:

Individual-level analysis aims to explore the attitudes of specific types of individuals. Two groups of individuals will be identified: Opinion leader, and individuals with certain ethnicities.

4.4.1 Opinion of Leaders

Opinion leaders on social media are active users who influence the opinions and behaviours of others and have access to wide social networks (Khan et al. 2015). Identifying opinion leaders can be crucial to promote behaviour change or to identify subjects that are of high interest to people. Identifying the attitudes of opinion leaders towards the Palestinian issue can be important because they reflect the attitudes of large sectors in their communities. In addition, identifying opinion leaders can help our politicians and decision makers to better promote for the Palestinian issue by, for example, supporting opinion leaders who stand by Palestine, and confront those who oppose the Palestinian rights.

Different metrics haven been used in the literature to identify opinion leaders on social networks. Some of these metrics have utilized the number of followers, interactions and activity, the leadership, or social network analysis (Ma & Liu 2014), (Zhai, Xu, & Jia 2008), and (Li, Ma, Zhang, & Huang 2013). In this work, opinion leaders will be identified by the number of followers so that users who have the most number of followers in a country will be treated as opinion leaders. Although it is not a perfect metric, but it is sufficient for the scope of this study.

We used the method proposed by Moore and McCabe (Moore et al. 1989) to identify users with extreme number of followers in each country. Moore and McCabe's method has been widely used in data analysis to find outliers in a distribution, whereas an outlier is the number that is more than 1.5 times the length of the box away from either the lower or upper quartiles. Refer to Section 2.2.6 for more information about the Moore and McCabe's method.

From a total of 38,328 users, 1,794 users were identified as opinion leaders. United States, Canada, and United Kingdom have the most number of opinion leaders, i.e. 59.14% of identified opinion leaders were from these countries. Table (4.6) shows statistics about the opinion leaders, while Table (4.7) shows top ten countries in terms of the number of opinion leaders.

Table (4.6): Opinion leaders' statistics

Avg. No. of followers per opinion leader	203623.49
St. Dev. of followers per opinion leader	89015.57
Avg. No. of tweets per opinion leader	13.76

Table (4.7): Top 10 countries with top number of leaders

No.	Country	No. of opinion leaders
1	United States	425
2	Canada	391
3	United Kingdom	299
4	France	95
5	India	61
6	Pakistan	35
7	Ecuador	31
8	Australia	29
9	Netherlands	23
10	South Africa	22

For some countries, no opinion leaders were identified, hence these countries were neglected from our analysis. Identified leaders were mostly official organizations, such as a newspaper, government officials, or media personnel. For example, the top opinion leaders in the US were Reuters Top News, Bernie Sanders and billboard, while top opinion leaders in the UK were The Economist, ABC News, and United Nations.

After identifying opinion leaders, the average friendliness scores for opinion leaders per country were calculated. A friendliness of an opinion leader is calculated using equation (4.1), which was used to calculate the country's friendliness, but by using only the tweets tweeted by the leader. We also calculated the standard deviation per country to identify the variance in friendliness of opinion leaders.

Figure (4.9) shows results for the top twenty countries in terms of friendliness of opinion leaders, while Figure (4.10) shows the standard deviation values of opinion leaders. The complete results can be found in Appendix F.

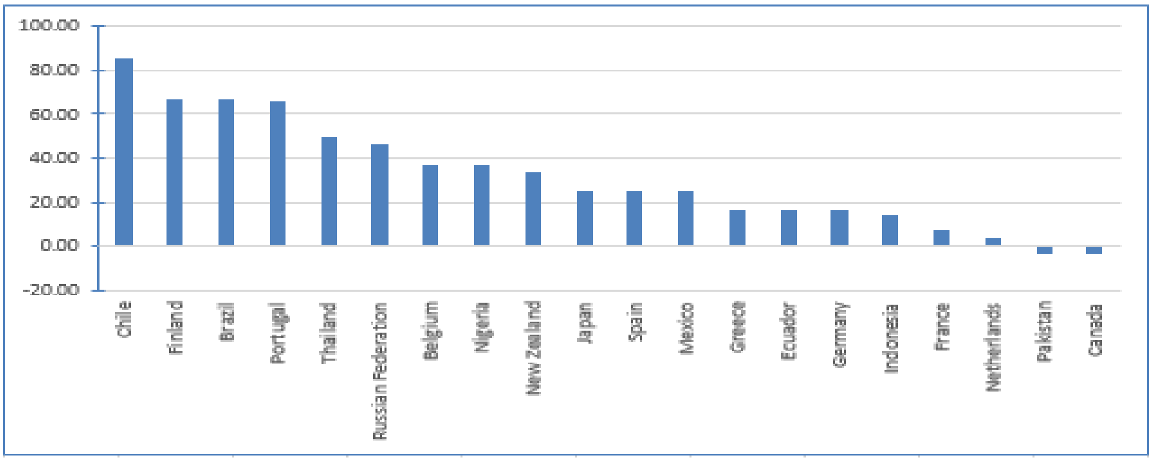

Figure (4.9): Leaders' friendliness

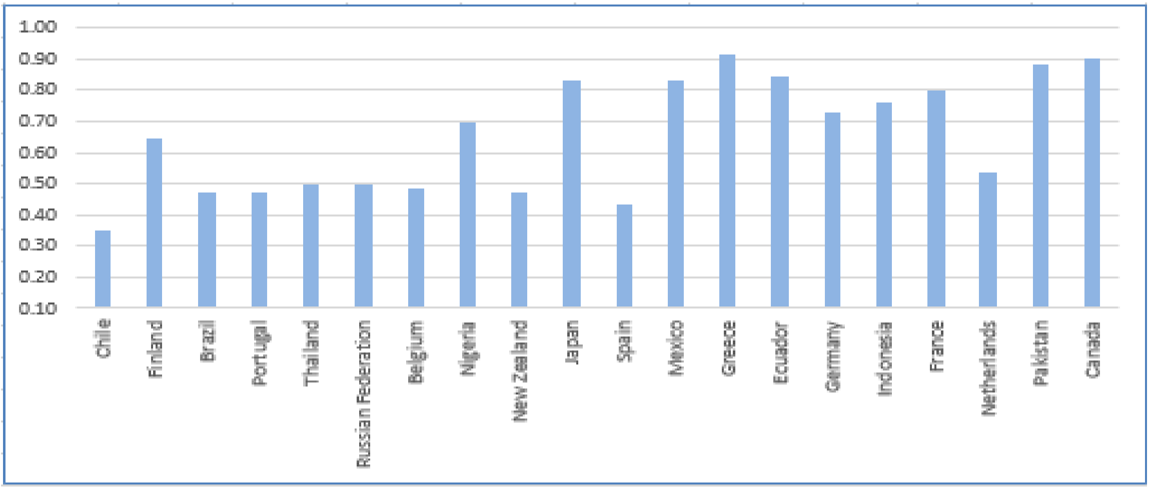

Figure (4.10): Leaders' standard dev.

Results show that opinion leaders from Chile, Finland and Brazil have the most favourable views of Palestine, while the leaders in Sweden, Islamic Republic of Iran and Kenya have the least favourable views. Looking at the standard deviations in Figure (4.10), it is noticeable that the divergence among opinion leaders increases where the friendliness scores of leaders are low, and vice versa. For example, the variance is high in countries like Pakistan and Canada in which the friendliness scores are low, while the variance is low in Chile and Spain.

Figure (4.11) show the friendliness scores of opinion leaders as compared to the friendliness of the top twenty countries that generate Palestine-focused tweets. In general, opinion leaders in most countries have more positive attitudes towards Palestine as compared to the attitude of the public opinion as in the cases of the UK, Brazil and Chile. However, countries like Japan, France and China have leaders with less favourable views as compared to the country's friendliness score.

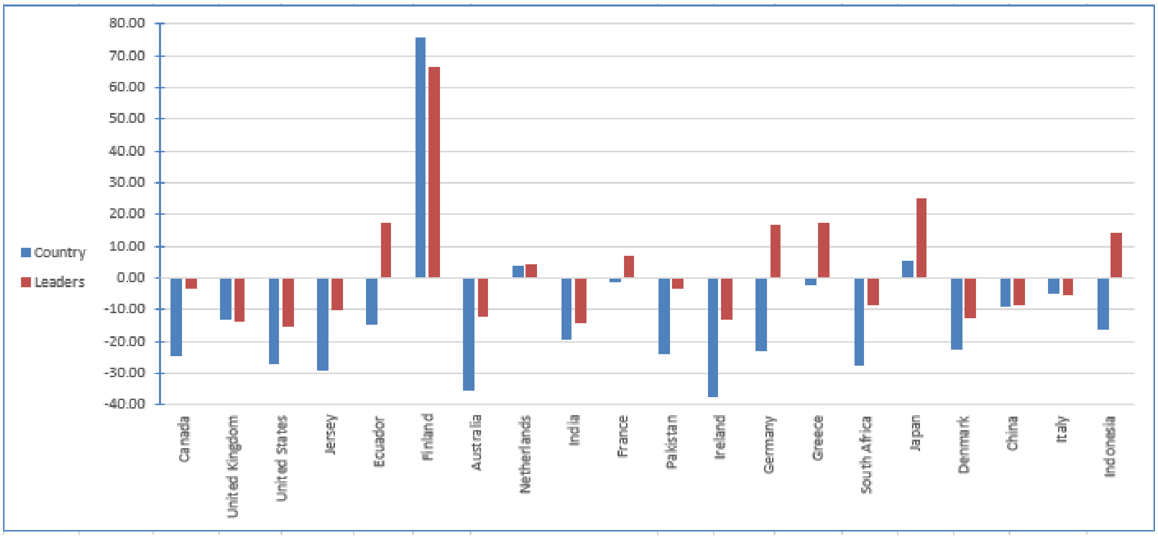

Figure (4.11): Friendliness of country and leaders

Table (4.8) shows details about sample opinion leaders.

Table (4.8): Sample about opinion leaders

Name	**Bio**	**Followers**	**Country**	**No. of tweets**	**Friendliness**	**St. Dev.**
Ronan Tynan	Filmmaker. Co-Founder with Anne Daly of award winning Esperanza Productions	20314	United Kingdom	353	65.72	0.94
Friends of Al Aqsa	UK based non-profit making NGO concerned with defending the human rights of Palestinians & protecting the sacred al-Aqsa Sanctuary in Jerusalem RT	18869	United Kingdom	255	75.49	0.86
Abbs Winston		25675	Canada	250	78.20	0.80
Palestine Video	A Palestine Vlog. Activist and other videos on Palestine. Proponent of One Democratic State! Definitely Palestinian	29906	Canada	209	91.39	0.56
Joe Catron	90% autotweets; 10% snark.	33562	Canada	193	55.70	0.92
Electronic Intifada	Palestine’s weapon of mass instruction.	81213	Canada	157	30.25	0.89
Oximity News Blend	Oximity bypasses all the filters of traditional news media, and delivers substantive news from organizations & individuals directly to readers around the world.	50960	United Kingdom	120	19.17	0.77
On security/Lundin	SIPRI Associate Fellow. Trustee Saferworld. Former EU Amb. Author EU and Security. Member Royal Academy of War Sciences	47924	Austria	111	35.59	0.92

Name	Bio	Followers	Country	No. of tweets	Friendliness	St. Dev.
The IMEU	The Institute for Middle East Understanding offers journalists quick access to information about Palestine	22094	Canada	109	53.67	0.95
Middle East Eye	Your eye on the ground. Independent Middle East and North Africa news	86923	Portugal	104	36.06	0.96

4.4.2 The Influence of Ethnicity

The term 'ethnicity' refers to a social group bound together by a more or less shared sense of historical (and sometimes geographical) origins which may be based upon language, culture, or religion(1). Members of the same ethnic group are likely to be sympathetic to each other's issues. Nowadays, a large number of Muslim and Arab people live in Europe and North America, and they have effective role in driving the economic, social and political aspects in these countries. In addition, a lot of Muslim and Arab organizations and individuals in these countries also provide continuous support to the Palestinian people. Part of this support comes through social networks in different forms such as retweet campaigns, hashtags, fundraising and promoted tweets.

In Section 4.3.2, the friendliness score of each country, which indicates the country's attitude towards Palestine, was measured. However, it is not clear to what extent the Muslim and Arab ethnic groups contribute to this attitude. The positive attitudes of Muslim and Arab ethnic groups are largely driven by the shared cultural and religious motivates. In contrast, non-Muslim groups may be motivated by different reasons such as human rights, humanitarian conditions, political or historical circumstances. Therefore, identifying the attitudes of different ethnic groups will be helpful for decision makers and social media activists to alter their speech and dialogue according to the needs and motivates of each ethnic group.

For simplicity, we decided to classify users from each country into two ethnic groups: people with Arabic names and people with non-Arabic names. Twitter's usernames were used to carry out this classification. People with Arabic names can include many people from other regions within the larger Muslim world such as Pakistanis, Iranians and Indians. Although it is not a perfect metric, the usernames can give an indicator to the ethnic group

(1) http://sites.cardiff.ac.uk/islamukcentre/rera/online-teaching-resources/muslims-in-britain-online-course/module-3-communities/diversities/muslim-ethnicities/

to which a Twitter user belongs. However, the limitation of using usernames is that some Twitter users may use nicknames not related to their original names.

To identify names, we used the website https://www.behindthename.com/ to get a comprehensive list of Arabic names written in English and Arabic. This website provides lists of common names for different ethnic groups. For Arabic, it includes 828 Arabic names in total, besides the different ways of writing these names in English. The list of Arabic names in both English and Arabic were extracted from the webpages. Afterwards, user names in our dataset were compared with the extracted names. If the username contains any Arabic name, in either Arabic or English language, it is added to the group of people of Arabic names.

To do this, we have used Jaro-Winkler distance algorithm(1) which is a string metric for measuring the edit distance between two sequences. Table (4.9) shows the sample results of the matching process. In total, 1,413 usernames were recognized as having Arabic names, while 38,712 usernames had no Arabic names.

Table (4.9): Results of matching Arabic Name

Arabic Name	User Name	Similarity score	Arabic Name	User Name	Similarity score
MOHAMED	Mahamed	0.94	KHALIFA	Amr Khalifa	0.91
ABBAS	Abbas Hussain	0.90	MAHDI	Mahdi Attar	0.90
ABD ALLAH	ABDALLAH	0.97	MUNIRA	Munir	0.93
ABDUL-AZIZ	Aziz	0.90	NABIL	Nabil H	0.92
ABUL-FAZL	Fazl	0.91	NABILA	Nabila Ramdani	0.90
ADIL	Adil Momin	0.90	OSAMA	Osama Bilal	0.90
AFIF	Afif Zet	0.90	RAFIQ	Arif Rafiq	0.90
AMINAH	Amina	0.93	USAMA	Usama Hasan	0.91
AS'AD	Asad Ali	0.92	USMAN	Usman Shehzad	0.90
BASSAM	MD Bassam	0.91	YASIN	Sara Yasin	0.91
DAUD	Moussa Daud	0.90	YASMIN	yasmina	0.94
DAWUD	Dawud Walid	0.91	ZIYAD	Ziyad – زياد	0.90
FARID	Kamal Farid	0.90	HAYDER	Haydeer	0.93
FARIHA	Fariha Fatima	0.91	MAHMOOD	Ze. Mahmood	0.90
HOUDA	Adam Houda	0.92	ABBAS	Abbas Hussain	0.91

(1) https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance/

After classifying users in each country into two groups, the friendliness of each group was measure, which is the average of friendliness scores of all members in the group.

Table (4.10) shows statistics about tweets generated by Arab and non-Arab groups in the top twenty countries in terms of Palestine-focused tweets. It also shows the average friendliness scores for the two groups. Overall, the percentage of tweets assigned to the Arab group is much smaller than the tweets assigned to the non-Arab group. This result is expected considering the fact that Arabs are considered a minority in Western countries. The complete results can be found in Appendix G.

Table (4.10): Arabic and Non-Arabic Friendliness

Country List			Arabic Ethnicity					Non-Arabic Ethnicity				
Name	**Focused Tweet**	**Friendliness**	**No. of tweets**	**Pos**	**Neg**	**Neu**	**Friendliness**	**No. of tweets**	**Pos**	**Neg**	**Neu**	**Friendliness**
Canada	27,490	**-24.43**	788	538	138	112	**50.76**	26,702	6,910	14,026	5,766	**-26.65**
United Kingdom	23,010	**-13.31**	389	210	103	76	**27.51**	22,621	8,071	11,241	3,309	**-14.01**
United States	20,125	**-27.04**	510	288	112	110	**34.51**	19,615	4,474	10,091	5,050	**-28.64**
Jersey	11,739	**-29.16**	12	8	1	3	**58.33**	11,727	3,420	6,850	1,457	**-29.25**
Ecuador	9,342	**-14.88**	42	14	11	17	**7.14**	9,300	358	1,751	7,191	**-14.87**
Finland	3,654	**75.97**	5	3	2	0	**20.00**	3,649	3,174	399	76	**76.05**
Australia	3,125	**-35.46**	53	26	14	13	**22.64**	3,072	728	1,848	496	**-36.46**
Netherlands	2,646	**3.82**	28	13	11	4	**7.14**	2,618	1,169	1,070	379	**3.78**
India	1,445	**-19.38**	91	24	23	44	**1.10**	1,354	272	553	529	**-20.75**
France	1,215	**-1.40**	26	12	6	8	**23.08**	1,189	428	451	310	**-1.93**
Pakistan	971	**-23.89**	209	113	57	39	**26.79**	762	138	426	198	**-37.80**
Ireland	971	**-37.38**	8	3	2	3	**12.50**	963	227	591	145	**-37.80**
Germany	830	**-23.01**	30	12	11	7	**3.33**	800	213	405	182	**-24.00**
Greece	820	**-2.56**	21	10	8	3	**9.52**	799	307	330	162	**-2.88**
South Africa	717	**-27.75**	37	16	17	4	**-2.70**	680	155	353	172	**-29.12**
Japan	642	**5.61**	2	0	0	2	**0.00**	640	308	272	60	**5.63**
Denmark	639	**-22.69**	22	4	3	15	**4.55**	617	173	319	125	**-23.66**
China	636	**-9.12**	97	38	35	24	**3.09**	539	153	214	172	**-11.32**
Italy	577	**-4.85**	15	8	5	2	**20.00**	562	199	230	133	**-5.52**
Indonesia	521	**-16.31**	14	4	2	8	**14.29**	507	145	232	130	**-17.16**

To understand the contribution of each ethnic group to the overall attitude of their countries, Figure (4.12) depicts the friendliness scores for the top twenty countries in terms of Palestine – focused tweets, while Figure (4.13) and Figure (4.14) shows the friendliness scores for Arab and non-Arab ethnicity respectively for the same countries. In general, Arab ethnicities in France, Jersey, and Japan have the highest influence on the country's attitude towards Palestine.

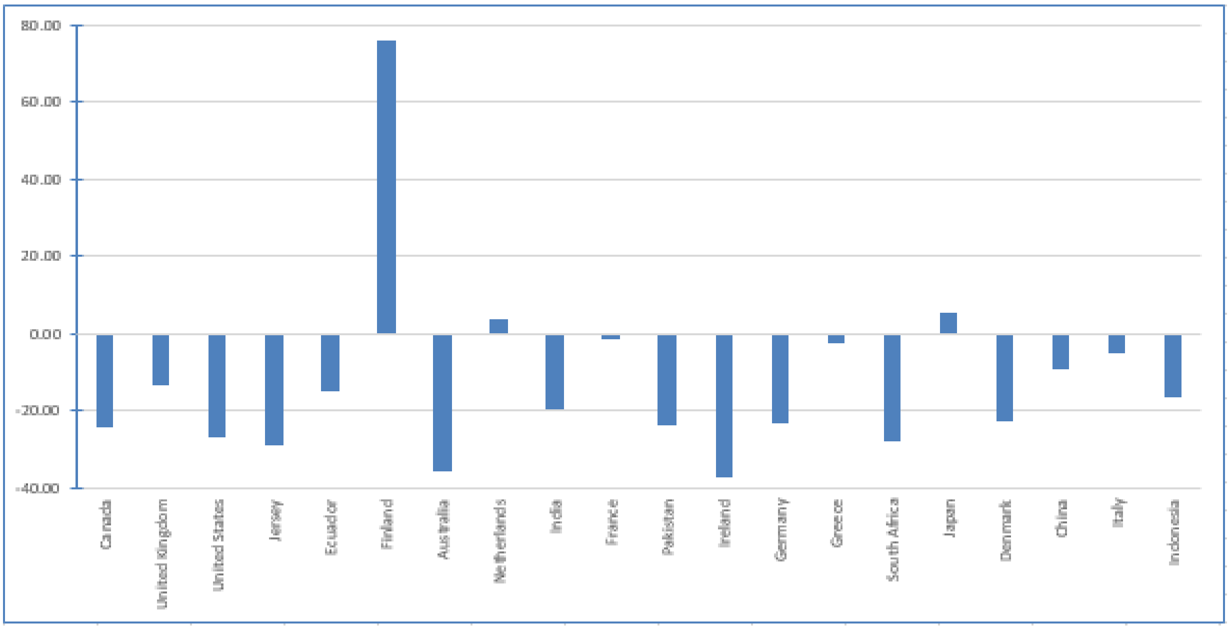

Figure (4.12): Country friendliness

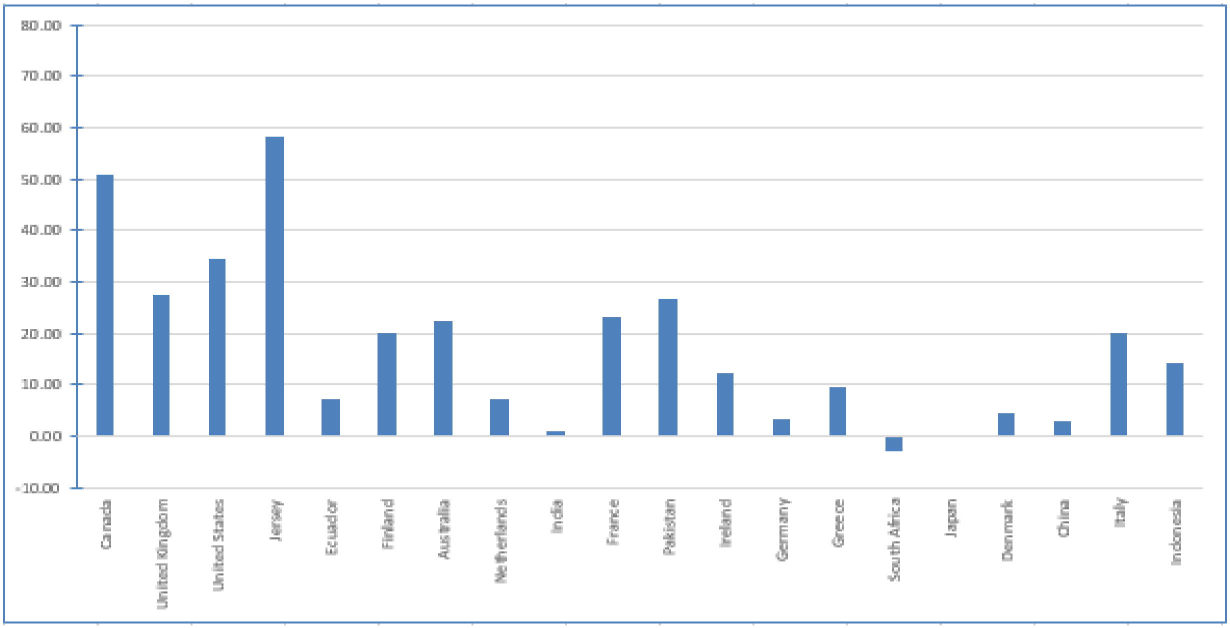

Figure (4.13): Arab ethnicity friendliness

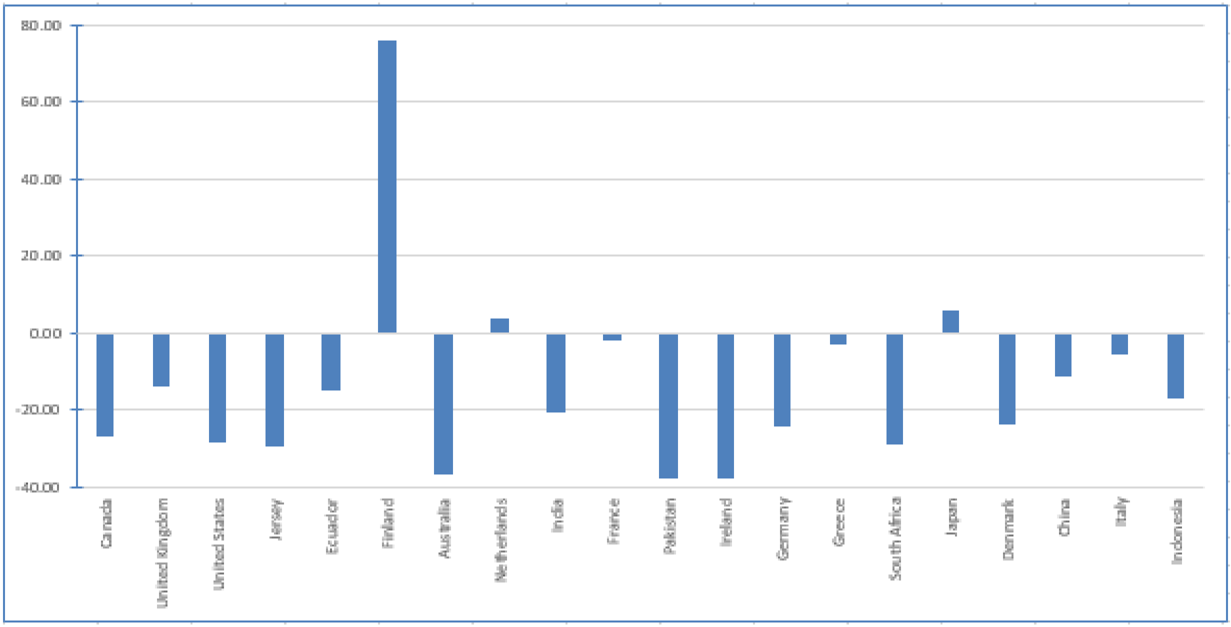

Figure (4.14): Non-Arab ethnicity friendliness

In general, results show that Arabic ethnic groups have more favourable attitudes towards Palestine than other groups. Apart from the case of Finland, non-Arabic groups have negative friendliness scores. The friendliness scores of Arabic ethnicities are high in Western countries such as Canada, the UK and the US where Muslim and Arab communities are large. However, they have not caused significant changes to the overall attitudes of their countries due to the small percentage of pertaining tweets.

Chapter 5

Conclusions

Chapter 5
Conclusions

This research aims to explore the attitudes towards Palestine in social media based on Twitter dataset. A dataset consisting of 178,524 tweets related to Palestine, and most tweeted during 2016, were collected and analyzed. The polarity of tweets was first determined by using a sentiment analyzer that was trained to identify the political attitude towards Palestine. The validity of the sentiment classifier to measure sentiments towards Palestine was also assessed. Afterwards, data analysis was performed at two levels: country-level analysis and individual-level analysis. The country-level analysis aimed to explore the country's overall interest in and attitude towards Palestine by: 1) Identifying counties that tweet most about Palestine. 2) Calculating friendliness scores for each country towards Palestine. 3) Analysing time series data to investigate how the country's attitude towards Palestine changes over time.

Results have shown that superpower countries such as the US, the UK and the Canada were among countries that generate most Palestine-focused tweets. However, they were not on the top in terms of friendliness. Countries like Finland, Brazil and Thailand were the friendliest considering the polarity of generated tweets. In general, the majority of countries had negative friendliness scores, and high divergence in opinion. This divergence was obvious in many European countries in particular such as France, the UK and Italy. This divergence can be seen as a positive point taking into consideration the Zionist control of the media, and the weak representation of pro-Palestinians in these countries.

Time-based analysis was also carried out for certain countries, and results showed that the attitude towards Palestine changes based on events that took place during the duration under study. These changes in attitude were explained by linking it with related events.

The individual-level analysis aimed to analyse data based on the activity of individuals. We first explored the attitudes of opinion leaders towards Palestine, and how they contribution to the overall attitude of their countries. Results have shown that opinion leaders in most countries had favourable views of Palestine, and that they were friendlier when compared to the friendliness scores of their countries. In addition, there was a weak correlation between the attitudes of opinion leaders and the overall attitudes of countries.

We also explored the influence of ethnicity on the public opinion, and the potential relation between the ethnicity of users and their attitudes towards Palestine. Results showed that Arab users have more positive attitude towards Palestine than non-Arabs. However, they have not caused significant changes to the overall attitudes of their countries.

There are many directions to extend this work in the future: First, we aim to use a larger dataset of tweets that span over several years. This will be likely to generate more reliable and generalizable results. Second, we aim to improve the sentiment analyser by training it with a larger volume of tweets. This is a critical point because the whole analysis is based on the polarities generated by the sentiment analyser. Third, we aim to explore and use more reliable approaches to identify opinion leaders and ethnic groups.

We think that other researchers, not necessarily from the IT discipline, can also build on these results to gain more insights: For example, results from this analysis may be compared with the results of related national polls in order to explore similarities and/or differences. Researchers may also study the attitude of each country and try to explain the rationale behind the perception of each country, its opinion leaders and ethnic groups towards Palestine.

References

References

Abdul-Mageed, Muhammad, Diab, Mona T, & Korayem, Mohammed. (2011). *Subjectivity and sentiment analysis of modern standard Arabic.* Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2.

Agarwal, Apoorv, Xie, Boyi, Vovsha, Ilia, Rambow, Owen, & Passonneau, Rebecca. (2011). *Sentiment analysis of twitter data.* Paper presented at the Proceedings of the workshop on languages in social media.

Aisopos, Fotis, Papadakis, George, Tserpes, Konstantinos, & Varvarigou, Theodora. (2012). *Content vs. context for sentiment analysis: a comparative analysis over microblogs.* Paper presented at the Proceedings of the 23rd ACM conference on Hypertext and social media.

Alias-i. (2008). LingPipe sentiment analyzer. Retrieved Dec 15, 2016, from http://alias-i.com/lingpipe

An, Xiaoran, Ganguly, Auroop R, Fang, Yi, Scyphers, Steven B, Hunter, Ann M, & Dy, Jennifer G. (2014). *Tracking climate change opinions from twitter data.* Paper presented at the Workshop on Data Science for Social Good.

Asur, Sitaram, & Huberman, Bernardo A. (2010). *Predicting the future with social media.* Paper presented at the Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on.

Baltas, Alexandros, Kanavos, Andreas, & Tsakalidis, Athanasios K. (2016). *An Apache Spark implementation for sentiment analysis on Twitter data.* Paper presented at the International Workshop of Algorithmic Aspects of Cloud Computing.

Barnaghi, Peiman, Ghaffari, Parsa, & Breslin, John G. (2016). *Opinion Mining and Sentiment Polarity on Twitter and Correlation Between Events and Sentiment.* Paper presented at the 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService).

Bermingham, Adam, & Smeaton, Alan F. (2011). On using Twitter to monitor political sentiment and predict election results.

Bollen, Johan, Mao, Huina, & Zeng, Xiaojun. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.

Ceron, Andrea, Curini, Luigi, & Iacus, Stefano M. (2015). Using sentiment analysis to monitor electoral campaigns method matters—Evidence from the United States and Italy. *Social Science Computer Review, 33*(1), 3-20.

Chen, Z, Fu, A, & Tang, J. (2002). Detection of outliered Patterns. *Dept. of CSE, Chinese University of Hong Kong*.

Compton, Ryan, Jurgens, David, & Allen, David. (2014). *Geotagging one hundred million twitter accounts with total variation minimization.* Paper presented at the Big Data (Big Data), 2014 IEEE International Conference on.

Cristianini, Nello, & Shawe-Taylor, John. (2000). An introduction to support vector machines: Cambridge University Press Cambridge.

Davidov, Dmitry, Tsur, Oren, & Rappoport, Ari. (2010). *Enhanced sentiment learning using twitter hashtags and smileys.* Paper presented at the Proceedings of the 23rd international conference on computational linguistics: posters.

Di Fatta, Giuseppe, Reade, J James, Jaworska, Sylvia, & Nanda, Anupam. (2015). *Big Social Data and Political Sentiment: the Tweet Stream during the UK General Election 2015 Campaign.* Paper presented at the 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity).

Ding, Xiaowen, Liu, Bing, & Yu, Philip S. (2008). *A holistic lexicon-based approach to opinion mining.* Paper presented at the Proceedings of the 2008 international conference on web search and data mining.

DNOiSE. (2017). Followthehashtag // Free twitter search analytics and business intelligence tool. Retrieved Dec 20, 2016, from http://www.followthehashtag.com/

Gimpel, Kevin, Schneider, Nathan, & O'Connor, Brendan. (2013). Annotation Guidelines for Twitter Part-of-Speech Tagging Version0.3 (March2013) http://www.ark.cs.cmu.edu/TweetNLP/annot_guidelines.pdf.

Go, Alec, Bhayani, Richa, & Huang, Lei. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford, 1*(12).

Godbole, Namrata, Srinivasaiah, Manja, & Skiena, Steven. (2007). Large-Scale Sentiment Analysis for News and Blogs. *ICWSM, 7*(21), 219-222.

Hernández, Sergio, & Sallis, Philip. (2011). *Sentiment-preserving reduction for social media analysis.* Paper presented at the Iberoamerican Congress on Pattern Recognition.

Hodeghatta, Umesh Rao. (2013). *Sentiment analysis of hollywood movies on twitter.* Paper presented at the Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

Hu, Minqing, & Liu, Bing. (2004). *Mining and summarizing customer reviews.* Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.

Ibrahim, Mochamad, Abdillah, Omar, Wicaksono, Alfan F, & Adriani, Mirna. (2015). *Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation.* Paper presented at the 2015 IEEE International Conference on Data Mining Workshop (ICDMW).

IDZELIS, M. Jazzy. (2005). The java open source spell checker. Retrieved Feb 16, 2017, from http://jazzy.sourceforge.net/

Jose, Rincy, & Chooralil, Varghese S. (2015). *Prediction of election result by enhanced sentiment analysis on Twitter data using Word Sense Disambiguation.* Paper presented at the 2015 International Conference on Control Communication & Computing India (ICCC).

Kantardzic, Mehmed, & Press, I. (2000). *Data mining: concepts, models, methods, and algorithms. 2003.* Paper presented at the Jiawei Han, Jian Pei, Yiwen Yin. Mining frequent patterns without candidate generation. SIGMOD'00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data.

Karau, Holden, Konwinski, Andy, Wendell, Patrick, & Zaharia, Matei. (2015). *Learning spark: lightning-fast big data analysis*: " O'Reilly Media, Inc.".

Kechaou, Zied, Ammar, Mohamed Ben, & Alimi, Adel M. (2011). *Improving e-learning with sentiment analysis of users' opinions.* Paper presented at the Global Engineering Education Conference (EDUCON), 2011 IEEE.

Khan, Nida Saddaf, Ata, Maira, & Rajput, Quratulain. (2015). *Identification of opinion leaders in social network.* Paper presented at the 2015 International Conference on Information and Communication Technologies (ICICT).

Li, Yanyan, Ma, Shaoqian, Zhang, Yonghe, & Huang, Ronghuai. (2013). An improved mix framework for opinion leader identification in online learning communities. *Knowledge-Based Systems, 43*, 43-51.

Lin, Dekang. (1998). *Automatic retrieval and clustering of similar words.* Paper presented at the Proceedings of the 17th international conference on Computational linguistics-Volume 2.

Liu, Bing. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.

Lohmann, Steffen, Burch, Michael, Schmauder, Hansjörg, & Weiskopf, Daniel. (2012). *Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds.* Paper presented at the Proceedings of the International Working Conference on Advanced Visual Interfaces.

Ma, Ning, & Liu, Yijun. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. *Expert Systems with Applications, 41*(4), 1357-1368.

Manning, Christopher D, Surdeanu, Mihai, Bauer, John, Finkel, Jenny Rose, Bethard, Steven, & McClosky, David. (2014). *The Stanford CoreNLP Natural Language Processing Toolkit.* Paper presented at the ACL (System Demonstrations).

Maynard, Diana, & Funk, Adam. (2011). *Automatic detection of political opinions in tweets.* Paper presented at the Extended Semantic Web Conference.

Maynard, Diana, Tablan, Valentin, Cunningham, Hamish, Ursu, Cristian, Saggion, Horacio, Bontcheva, Kalina, & Wilks, Yorick. (2002). Architectural elements of language engineering robustness. *Natural Language Engineering, 8*(2-3), 257-274.

Meng, Xiangrui, Bradley, Joseph, Yavuz, Burak, Sparks, Evan, Venkataraman, Shivaram, Liu, Davies, . . . Owen, Sean. (2016). Mllib: Machine learning in apache spark. *The Journal of Machine Learning Research, 17*(1), 1235-1241.

Montesinos, Lucas, Rodr, S Juan Pablo, Orchard, Marcos, & Eyheramendy, Susana. (2015). *Sentiment analysis and prediction of events in TWITTER.* Paper presented at the 2015

CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON).

Moore, David S, & McCabe, George P. (1989). *Introduction to the Practice of Statistics*: WH Freeman/Times Books/Henry Holt & Co.

Nodarakis, Nikolaos, Sioutas, Spyros, Tsakalidis, Athanasios K, & Tzimas, Giannis. (2016). *Large Scale Sentiment Analysis on Twitter with Spark.* Paper presented at the EDBT/ICDT Workshops.

O'Connor, Brendan, Balasubramanyan, Ramnath, Routledge, Bryan R, & Smith, Noah A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM, 11*(122-129), 1-2.

Olson, David L, & Delen, Dursun. (2008). *Advanced data mining techniques*: Springer Science & Business Media.

Osimo, David, & Mureddu, Francesco. (2012). Research challenge on opinion mining and sentiment analysis. *Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment, 508*.

Owoputi, Olutobi, O'Connor, Brendan, Dyer, Chris, Gimpel, Kevin, Schneider, Nathan, & Smith, Noah A. (2013). *Improved part-of-speech tagging for online conversational text with word clusters*.

Pak, Alexander, & Paroubek, Patrick. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining.* Paper presented at the LREc.

Pang, Bo, Lee, Lillian, & Vaithyanathan, Shivakumar. (2002). *Thumbs up?: sentiment classification using machine learning techniques.* Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.

Paul, Michael J, & Dredze, Mark. (2011). You are what you Tweet: Analyzing Twitter for public health.

Qaisi, Laila M, & Aljarah, Ibrahim. (2016). A Twitter Sentiment Analysis for Cloud Providers: A Case Study of Azure vs. A WS.

Ramaswamy, Sridhar, Rastogi, Rajeev, & Shim, Kyuseok. (2000). *Efficient algorithms for mining outliers from large data sets.* Paper presented at the ACM Sigmod Record.

Riloff, Ellen, Patwardhan, Siddharth, & Wiebe, Janyce. (2006). *Feature subsumption for opinion analysis.* Paper presented at the Proceedings of the 2006 conference on empirical methods in natural language processing.

Roebuck, Kevin. (2011). Sentiment Analysis (ebook). Emereo Pty Limited. 1-2.

Salah, Z, Coenen, F, & Grossi, D. (2014). Political sentiment analysis: Predicting speaker attitude in the UK House of.

Sane, Shirish S, & Ghatol, Ashok A. (2006). *Use of instance typicality for efficient detection of outliers with neural network classifiers.* Paper presented at the Information Technology, 2006. ICIT'06. 9th International Conference on.

Sarlan, Aliza, Nadam, Chayanit, & Basri, Shuib. (2014). *Twitter sentiment analysis.* Paper presented at the Information Technology and Multimedia (ICIMU), 2014 International Conference on.

Shukri, Sarah E, Yaghi, Rawan I, Aljarah, Ibrahim, & Alsawalqah, Hamad. (2015). *Twitter sentiment analysis: A case study in the automotive industry.* Paper presented at the Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on.

Taboada, Maite, Brooke, Julian, Tofiloski, Milan, Voll, Kimberly, & Stede, Manfred. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics, 37*(2), 267-307.

Thelwall, Mike, Buckley, Kevan, & Paltoglou, Georgios. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology, 63*(1), 163-173.

Tumasjan, Andranik, Sprenger, Timm Oliver, Sandner, Philipp G, & Welpe, Isabell M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM, 10*, 178-185.

Turney, Peter D. (2002). *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews.* Paper presented at the Proceedings of the 40th annual meeting on association for computational linguistics.

Twitter, Inc. (2017). Twitter Developer Documentation. Retrieved Aug 2, 2017, from https://dev.twitter.com/streaming/overview

Whissell, CM. (1989). The dictionary of affect in language. emotion: theory, research and experience 4. The Measurement of Emotions: Academic Press, New York.

Wiebe, Janyce. (2000). *Learning subjective adjectives from corpora.* Paper presented at the AAAI/IAAI.

Wiebe, Janyce M, Bruce, Rebecca F, & O'Hara, Thomas P. (1999). *Development and use of a gold-standard data set for subjectivity classifications.* Paper presented at the Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics.

Wiebe, Janyce, & Riloff, Ellen. (2005). *Creating subjective and objective sentence classifiers from unannotated texts.* Paper presented at the International Conference on Intelligent Text Processing and Computational Linguistics.

Younus, Arjumand, Qureshi, M Atif, Asar, Fiza Fatima, Azam, Muhammad, Saeed, Muhammad, & Touheed, Nasir. (2011). *What do the average twitterers say: A twitter model for public opinion analysis in the face of major political events.* Paper presented at the Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on.

Yusuke, Yamamoto. (2007). Java library for the Twitter API. Retrieved Oct 20, 2016, from https://github.com/yusuke

Zaharia, Matei, Chowdhury, Mosharaf, Franklin, Michael J, Shenker, Scott, & Stoica, Ion. (2010). Spark: Cluster computing with working sets. *HotCloud, 10*(10-10), 95.

Zhai, Zhongwu, Xu, Hua, & Jia, Peifa. (2008). *Identifying opinion leaders in BBS.* Paper presented at the Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 03.

Zhou, Xujuan, Tao, Xiaohui, Yong, Jianming, & Yang, Zhenyu. (2013). *Sentiment analysis on tweets for social events.* Paper presented at the Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on.

Appendix A

Date	User Name	Nickname	Tweet content	Favs	RTs	Latitude	Longitude	Country	Place (as appears on Bio)	Followers	Following	Tweet language	Is a RT	Hashtags
12/19/2016	Middle East Monitor	MiddleEastMnt	Israel bans burials in parts of East #Jerusalem Muslim cemetery. https://t.co/irQV7AePpd | #Palestine https://t.co/HIcMlaVkzB	1	6	51.509	-0.126	GB	London, UK	54083	982	En	FALSE	#Jerusalem,#Palestine
12/19/2016	Robby Martin	takethepss	#Palestine Post-truth and lies have always been Israel's way, and the West plays its role https://t.co/AwYuGhkzCt via @MiddleEastMnt			53.33306	-6.24889	IE	Ireland	2779	2361	En	FALSE	#Palestine
12/19/2016	ISRAEL BOMBS BABIES	Col_Connaughton	Gaza's ambulances under attack https://t.co/OU9emF3ZQs #palestine #israel #gaza #BDS			51.509	-0.126	GB	London UK	4693	2028	En	FALSE	#palestine,#israel,#gaza,#BDS
12/19/2016	ISRAEL BOMBS BABIES	Col_Connaughton	Gaza's hospitals on the frontlines https://t.co/bpsdXBYA5Z #palestine #israel #gaza #BDS			51.509	-0.126	GB	London UK	4693	2028	En	FALSE	#palestine,#israel,#gaza,#BDS
12/19/2016	ISRAEL BOMBS BABIES	Col_Connaughton	Obstructions to Medical Access in Gaza: Bader's Story https://t.co/vB4xNe8hgA #palestine #israel #gaza #BDS			51.509	-0.126	GB	London UK	4693	2028	En	FALSE	#palestine,#israel,#gaza,#BDS
12/19/2016	Al Salam	queenrose0	STOP Israel's terrorism against children!! @aymansayehh @Palestine_on @Abbas_q88t @bailly_brigitte @Free_Palstine… https://t.co/MX0ANhigi7							189	380	En	FALSE	
12/19/2016	Diana Greenwald	hispeedtourist	What happens when one part of your state urges you to do something that another part of your state deems illegal…. https://t.co/yLqs28Cb7U							605	1162	En	FALSE	
12/19/2016	Palestine Trends	Palestinolizer	Mustafa Barghouti: Trump Pick for Ambassador to Israel Giving the Green Light to... https://t.co/Qo3t6qcG9B https://t.co/EyYFDW9kBt			31.5	34.46667	PS	Gaza City	2536	1147	En	FALSE	
12/19/2016	Joe Iosbaker	iosbakerjoe	https://t.co/ulIvo4lSIN Trump anti US in Iraq, Syria - but pro occupation of Palestine. BS Artist! Support for Israel = more US wars!			41.85003	-87.65	US	Chicago	798	481	En	FALSE	
12/19/2016	ISRAEL BOMBS BABIES	Col_Connaughton	911 solved. #gaza #palestine #israel #BDS #911 #mossad #insidejob https://t.co/CdyEq6iJS8		1	51.509	-0.126	GB	London UK	4693	2028	En	FALSE	#gaza,#palestine,#israel,#BDS
12/19/2016	Christoph Paul	ChristophPaul_	DJ Khaled Presents One-State Solution to Israel-Palestine Conflict: 'WeTheBest-istan' - The Mideast Beast https://t.co/zTS785NQ0C	1						18198	3694	En	FALSE	
12/19/2016	Charafa	freeworldun	Israel bans burials in parts of East Jerusalem Muslim cemetery #Palestine @Hrw https://t.co/M82pgLjuMk							12273	10830	En	FALSE	#Palestine

Date	User Name	Nickname	Tweet content	Favs	RTs	Latitude	Longitude	Country	Place (as appears on Bio)	Followers	Following	Tweet language	Is a RT	Hashtags
12/19/2016	Al Salam	queenrose0	@ex2Tory @RosmeWarda @aymansayehh @Palestine_on God didn't choose terrorists kills children and civilians daily! Bo… https://t.co/G8wodWITb2	1	1					189	380	En	FALSE	
12/19/2016	ISRAEL BOMBS BABIES	Col_Connaughton	How effective is the anti-Israel BDS movement? https://t.co/JYNMEOzggY #gaza #palestine #israel #BDS			51.509	-0.126	GB	London UK	4693	2028	En	FALSE	#gaza,#palestine,#israel,#BDS
12/19/2016	Charafa	freeworldun	Israel threatens to toss journalist Antony Loewenstein after he asked Lapid question about #apartheid in #Palestine https://t.co/Fg0AF17GY0	3	8					12273	10830	En	FALSE	#apartheid,#Palestine
12/19/2016	ıʇʇɐɥs-ɹɐ pǝlɐɥʞ	ShattiUnique	@BoykoBorissov i hope can you say the same about what #Israel doing to #Palestine .. and #Al-Asad doing to #Syria people							99	48	En	FALSE	#Israel,#Palestine,#Al,#Syria
12/19/2016	Susan John-Richards	SJRTooting	Israel BANS burials in East Jerusalem Muslim cemetery- https://t.co/mkrHxg0SQ3		2					10610	8715	En	FALSE	
12/19/2016	Palestine Trends	Palestinolizer	Trump Taps His Far-Right-Wing Bankruptcy Lawyer to be U.S. Ambassador to Israel https://t.co/GwejcpoA0s https://t.co/OccxxxjPe4			31.5	34.46667	PS	Gaza City	2536	1147	En	FALSE	
12/19/2016	ISRAEL BOMBS BABIES	Col_Connaughton	Pink Floyd - Song for Palestine: https://t.co/ozAq6Oov7u #gaza #palestine #israel #BDS			51.509	-0.126	GB	London UK	4693	2028	En	FALSE	#gaza,#palestine,#israel,#BDS

Appendix B

No.	Country	Country code	Focused Tweets
1	Canada	CA	27,490
2	United Kingdom	GB	23,010
3	United States	US	20,125
4	Jersey	JE	11,739
5	Ecuador	EC	9,342
6	Finland	FI	3,654
7	Australia	AU	3,125
8	Netherlands	NL	2,646
9	India	IN	1,445
10	France	FR	1,215
11	Ireland	IE	971
12	Pakistan	PK	971
13	Germany	DE	830
14	Greece	GR	820
15	South Africa	ZA	717
16	Japan	JP	642
17	Denmark	DK	639
18	China	CN	636
19	Italy	IT	577
20	Indonesia	ID	521
21	Spain	ES	422
22	Belgium	BE	399
23	Turkey	TR	383
24	Brazil	BR	382
25	Switzerland	CH	381
26	Kazakhstan	KZ	352
27	Nigeria	NG	315
28	Sweden	SE	293
29	Mexico	MX	278
30	Thailand	TH	262
31	Russian Federation	RU	257
32	Portugal	PT	249
33	Malaysia	MY	220
34	Islamic Republic of Iran	IR	218
35	Kenya	KE	216
36	Chile	CL	206
37	Slovenia	SI	179
38	New Zealand	NZ	177
39	Austria	AT	172

Appendix C

No.	Country	code	Focused Tweets	Population	focused-tweet by capita
1	Canada	CA	27,490	36,289,822	0.075751267
2	United Kingdom	GB	23,010	65,788,574	0.034975678
3	United States	US	20,125	322,179,605	0.006246516
4	Jersey	JE	11,739	164,541	7.134392036
5	Ecuador	EC	9,342	16,385,068	0.057015326
6	Finland	FI	3,654	5,503,132	0.066398553
7	Australia	AU	3,125	24,125,848	0.012952913
8	Netherlands	NL	2,646	16,987,330	0.015576315
9	India	IN	1,445	1,324,171,354	0.000109125
10	France	FR	1,215	64,720,690	0.001877298
11	Ireland	IE	971	4,726,078	0.020545577
12	Pakistan	PK	971	193,203,476	0.000502579
13	Germany	DE	830	81,914,672	0.001013249
14	Greece	GR	820	11,183,716	0.007332089
15	South Africa	ZA	717	56,015,473	0.001280003
16	Japan	JP	642	127,748,513	0.00050255
17	Denmark	DK	639	5,711,870	0.011187229
18	China	CN	636	1,403,500,365	4.53153E-05
19	Italy	IT	577	59,429,938	0.000970891
20	Indonesia	ID	521	261,115,456	0.000199529
21	Spain	ES	422	46,347,576	0.000910511
22	Belgium	BE	399	11,358,379	0.003512825
23	Turkey	TR	383	79,512,426	0.000481686
24	Brazil	BR	382	207,652,865	0.000183961
25	Switzerland	CH	381	8,401,739	0.004534775
26	Kazakhstan	KZ	352	17,987,736	0.001956889
27	Nigeria	NG	315	185,989,640	0.000169364
28	Sweden	SE	293	9,837,533	0.002978389
29	Mexico	MX	278	127,540,423	0.00021797
30	Thailand	TH	262	68,863,514	0.000380463
31	Russian Federation	RU	257	143,964,513	0.000178516
32	Portugal	PT	249	10,371,627	0.002400781
33	Malaysia	MY	220	31,187,265	0.000705416
34	Islamic Republic of Iran	IR	218	80,277,428	0.000271558
35	Kenya	KE	216	48,461,567	0.000445714
36	Chile	CL	206	17,909,754	0.001150211
37	Slovenia	SI	179	2,077,862	0.008614624
38	New Zealand	NZ	177	24,125,848	0.000733653
39	Austria	AT	172	8,712,137	0.001974257

Appendix D

No.	Country	Focused Tweets	No. of tweets contain *Palestine* keyword	Log	Google index 2016
1	Canada	27,490	17804	14.12	85
2	United Kingdom	23,010	17429	14.09	82
3	United States	20,125	12100	13.56	79
4	Jersey	11,739	11611	13.50	67
5	Ecuador	9,342	8967	13.13	64
6	Finland	3,654	2296	11.16	76
7	Australia	3,125	1289	10.33	68
8	Netherlands	2,646	1488	10.54	58
9	India	1,445	1120	10.13	60
10	France	1,215	911	9.83	63
11	Ireland	971	635	9.31	75
12	Pakistan	971	714	9.48	78
13	Germany	830	513	9.00	55
14	Greece	820	459	8.84	43
15	South Africa	717	441	8.78	59
16	Japan	642	384	8.58	70
17	Denmark	639	411	8.68	65
18	China	636	391	8.61	73
19	Italy	577	288	8.17	63
20	Indonesia	521	279	8.12	55
21	Spain	422	269	8.07	52
22	Belgium	399	327	8.35	60
23	Turkey	383	291	8.18	36
24	Brazil	382	208	7.70	42
25	Switzerland	381	284	8.15	30
26	Kazakhstan	352	351	8.46	20
27	Nigeria	315	221	7.79	43
28	Sweden	293	172	7.43	57
29	Mexico	278	169	7.40	52
30	Thailand	262	203	7.67	32
31	Russian Federation	257	184	7.52	73
32	Portugal	249	150	7.23	56
33	Malaysia	220	158	7.30	65
34	Islamic Republic of Iran	218	104	6.70	44
35	Kenya	216	164	7.36	43
36	Chile	206	122	6.93	65
37	New Zealand	177	135	7.08	66
38	Austria	172	152	7.25	52

Appendix E

No.	Country	Focused Tweets	Positive	Negative	Neutral	Friendliness	St. Dev
1	Finland	3,654	3,177	401	76	**75.97**	**0.63**
2	Brazil	382	184	118	80	**17.28**	**0.87**
3	Thailand	262	127	89	46	**14.50**	**0.90**
4	Japan	642	308	272	62	**5.61**	**0.95**
5	Netherlands	2,646	1,182	1,081	383	**3.82**	**0.92**
6	France	1,215	440	457	318	**-1.40**	**0.86**
7	Greece	820	317	338	165	**-2.56**	**0.89**
8	Nigeria	315	104	118	93	**-4.44**	**0.84**
9	Italy	577	207	235	135	**-4.85**	**0.87**
10	Islamic Republic of Iran	218	80	96	42	**-7.34**	**0.90**
11	Portugal	249	91	112	46	**-8.43**	**0.90**
12	China	636	191	249	196	**-9.12**	**0.83**
13	Chile	206	77	96	33	**-9.22**	**0.91**
14	United Kingdom	23,010	8,281	11,344	3,385	**-13.31**	**0.91**
15	Ecuador	9,342	372	1,762	7,208	**-14.88**	**0.45**
16	Indonesia	521	149	234	138	**-16.31**	**0.84**
17	Sweden	293	90	139	64	**-16.72**	**0.87**
18	Turkey	383	88	157	138	**-18.02**	**0.78**
19	Kenya	216	58	97	61	**-18.06**	**0.83**
20	India	1,445	296	576	573	**-19.38**	**0.75**
21	Slovenia	179	46	83	50	**-20.67**	**0.82**
22	Spain	422	102	196	124	**-22.27**	**0.81**
23	Denmark	639	177	322	140	**-22.69**	**0.85**
24	Germany	830	225	416	189	**-23.01**	**0.85**
25	Mexico	278	65	131	82	**-23.74**	**0.81**
26	Pakistan	971	251	483	237	**-23.89**	**0.84**
27	Belgium	399	90	186	123	**-24.06**	**0.80**
28	Canada	27,490	7,448	14,164	5,878	**-24.43**	**0.85**
29	New Zealand	177	44	88	45	**-24.86**	**0.83**
30	Russian Federation	257	60	125	72	**-25.29**	**0.81**
31	Malaysia	220	40	97	83	**-25.91**	**0.75**
32	United States	20,125	4,762	10,203	5,160	**-27.04**	**0.82**
33	South Africa	717	171	370	176	**-27.75**	**0.82**
34	Jersey	11,739	3,428	6,851	1,460	**-29.16**	**0.89**
35	Australia	3,125	754	1,862	508	**-35.46**	**0.84**
36	Ireland	971	230	593	148	**-37.38**	**0.84**
37	Switzerland	381	72	258	51	**-48.82**	**0.79**
38	Austria	172	15	134	23	**-69.19**	**0.62**
39	Kazakhstan	352	12	311	29	**-84.94**	**0.44**

Appendix F

No.	Country	Focused Tweets	Leaders' tweets	Positive	Negative	Neutral	Friendliness	St. Dev
1	Chile	206	21	18	0	3	**85.71**	**0.35**
2	Finland	3,654	21	16	2	3	**66.67**	**0.64**
3	Brazil	382	3	2	0	1	**66.67**	**0.47**
4	Portugal	249	50	33	0	17	**66.00**	**0.47**
5	Thailand	262	2	1	0	1	**50.00**	**0.50**
6	Russian Federation	257	52	24	0	28	**46.15**	**0.50**
7	Belgium	399	16	6	0	10	**37.50**	**0.48**
8	Nigeria	315	8	4	1	3	**37.50**	**0.70**
9	New Zealand	177	3	1	0	2	**33.33**	**0.47**
10	Japan	642	8	4	2	2	**25.00**	**0.83**
11	Spain	422	4	1	0	3	**25.00**	**0.43**
12	Mexico	278	4	2	1	1	**25.00**	**0.83**
13	Greece	820	29	15	10	4	**17.24**	**0.91**
14	Ecuador	9,342	70	32	20	18	**17.14**	**0.84**
15	Germany	830	36	13	7	16	**16.67**	**0.73**
16	Indonesia	521	35	13	8	14	**14.29**	**0.76**
17	France	1,215	28	10	8	10	**7.14**	**0.80**
18	Netherlands	2,646	89	15	11	63	**4.49**	**0.54**
19	Pakistan	971	59	22	24	13	**-3.39**	**0.88**
20	Canada	27,490	5,361	2,084	2,274	1,003	**-3.54**	**0.90**
21	Italy	577	38	8	10	20	**-5.26**	**0.69**
22	China	636	12	2	3	7	**-8.33**	**0.64**
23	South Africa	717	47	16	20	11	**-8.51**	**0.87**
24	Austria	172	111	1	11	99	**-9.01**	**0.32**
25	Jersey	11,739	260	52	78	130	**-10.00**	**0.70**
26	Australia	3,125	57	19	26	12	**-12.28**	**0.88**
27	Denmark	639	8	3	4	1	**-12.50**	**0.93**
28	Turkey	383	55	11	18	26	**-12.73**	**0.71**
29	Ireland	971	15	5	7	3	**-13.33**	**0.88**
30	United Kingdom	23,010	2,220	703	1,007	510	**-13.69**	**0.87**
31	India	1,445	142	31	51	60	**-14.08**	**0.75**
32	United States	20,125	1,412	338	552	522	**-15.16**	**0.78**
33	Switzerland	381	3	0	1	2	**-33.33**	**0.47**
34	Malaysia	220	14	0	8	6	**-57.14**	**0.49**
35	Sweden	293	4	0	3	1	**-75.00**	**0.43**
36	Islamic Republic of Iran	218	57	0	43	14	**-75.44**	**0.43**
37	Kenya	216	13	0	11	2	**-84.62**	**0.36**
38	Kazakhstan	352	1	0	1	0	**-100.00**	**0.00**
39	Slovenia	179	1	0	1	0	**-100.00**	**0.00**

Appendix G

Country List			Arabic Ethnicity					Non-Arabic Ethnicity				
Name	**Focused Tweet**	**Friendliness**	**No. of tweets**	**Pos**	**Neg**	**Neu**	**Friendliness**	**No. of tweets**	**Pos**	**Neg**	**Neu**	**Friendliness**
Canada	27,490	**-24.43**	788	538	138	112	**50.76**	26,702	6,910	14,026	5,766	**-26.65**
United Kingdom	23,010	**-13.31**	389	210	103	76	**27.51**	22,621	8,071	11,241	3,309	**-14.01**
United States	20,125	**-27.04**	510	288	112	110	**34.51**	19,615	4,474	10,091	5,050	**-28.64**
Jersey	11,739	**-29.16**	12	8	1	3	**58.33**	11,727	3,420	6,850	1,457	**-29.25**
Ecuador	9,342	**-14.88**	42	14	11	17	**7.14**	9,300	358	1,751	7,191	**-14.98**
Finland	3,654	**75.97**	5	3	2	0	**20.00**	3,649	3,174	399	76	**76.05**
Australia	3,125	**-35.46**	53	26	14	13	**22.64**	3,072	728	1,848	496	**-36.46**
Netherlands	2,646	**3.82**	28	13	11	4	**7.14**	2,618	1,169	1,070	379	**3.78**
India	1,445	**-19.38**	91	24	23	44	**1.10**	1,354	272	553	529	**-20.75**
France	1,215	**-1.40**	26	12	6	8	**23.08**	1,189	428	451	310	**-1.93**
Pakistan	971	**-23.89**	209	113	57	39	**26.79**	762	138	426	198	**-37.80**
Ireland	971	**-37.38**	8	3	2	3	**12.50**	963	227	591	145	**-37.80**
Germany	830	**-23.01**	30	12	11	7	**3.33**	800	213	405	182	**-24.00**
Greece	820	**-2.56**	21	10	8	3	**9.52**	799	307	330	162	**-2.88**
South Africa	717	**-27.75**	37	16	17	4	**-2.70**	680	155	353	172	**-29.12**
Japan	642	**5.61**	2	0	0	2	**0.00**	640	308	272	60	**5.63**
Denmark	639	**-22.69**	22	4	3	15	**4.55**	617	173	319	125	**-23.66**
China	636	**-9.12**	97	38	35	24	**3.09**	539	153	214	172	**-11.32**
Italy	577	**-4.85**	15	8	5	2	**20.00**	562	199	230	133	**-5.52**
Indonesia	521	**-16.31**	14	4	2	8	**14.29**	507	145	232	130	**-17.16**
Spain	422	**-22.27**	3	0	1	2	**-33.33**	419	102	195	122	**-22.20**
Belgium	399	**-24.06**	4	1	1	2	**0.00**	395	89	185	121	**-24.30**
Turkey	383	**-18.02**	11	6	3	2	**27.27**	372	82	154	136	**-19.35**
Brazil	382	**17.28**	7	3	2	2	**14.29**	375	181	116	78	**17.33**

Country List			Arabic Ethnicity					Non-Arabic Ethnicity				
Name	**Focused Tweet**	**Friendliness**	**No. of tweets**	**Pos**	**Neg**	**Neu**	**Friendliness**	**No. of tweets**	**Pos**	**Neg**	**Neu**	**Friendliness**
Switzerland	381	**-48.82**	3	0	1	2	**-33.33**	378	72	257	49	**-48.94**
Kazakhstan	352	**-84.94**	81	3	9	69	**-7.41**	271	9	302	-40	**-108.12**
Nigeria	315	**-4.44**	19	3	2	14	**5.26**	296	101	116	79	**-5.07**
Sweden	293	**-16.72**	4	2	1	1	**25.00**	289	88	138	63	**-17.30**
Mexico	278	**-23.74**	3	0	0	3	**0.00**	275	65	131	79	**-24.00**
Thailand	262	**14.50**	0	**0**	**0**	**0**	**0.00**	262	127	89	46	**14.50**
Russian Federation	257	**-25.29**	14	6	4	4	**14.29**	243	54	121	68	**-27.57**
Portugal	249	**-8.43**	10	4	5	1	**-10.00**	239	87	107	45	**-8.37**
Malaysia	220	**-25.91**	30	9	7	14	**6.67**	190	31	90	69	**-31.05**
Islamic Republic of Iran	218	**-7.34**	19	8	6	5	**10.53**	199	72	90	37	**-9.05**
Kenya	216	**-18.06**	20	8	5	7	**15.00**	196	50	92	54	**-21.43**
Chile	206	**-9.22**	1	1	0	0	**100.00**	205	76	96	33	**-9.76**
Slovenia	179	**-20.67**	9	4	3	2	**11.11**	170	42	80	48	**-22.35**
New Zealand	177	**-24.86**	3	0	0	3	**0.00**	174	44	88	42	**-25.29**
Austria	172	**-4.65**	4	3	1	0	**50.00**	168	12	22	134	**-5.95**

Appendix H

a	about	all	am	an	and	any	are
as	at	be	been	by	do	for	from
had	has	have	he	her	here	hers	him
his	i	if	in	into	is	it	its
me	more	most	my	of	off	on	or
other	she	so	some	such	that	the	there
these	they	this	those	to	too	up	was
we	which	while	who	you	your	yours	but
dose	once	only	then	with	will		