ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

SOCIAL MEDIA DATA VALUATION MODEL FOR DISASTER INCIDENCE MAPPING

Ph.D. THESIS

Ayse Giz GULNERMAN GENGEC

Department of Geomatics Engineering

Geomatics Engineering Programme

OCTOBER 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

SOCIAL MEDIA DATA VALUATION MODEL FOR DISASTER INCIDENCE MAPPING

Ph.D.THESIS

Ayse Giz GULNERMAN GENGEC
(501142609)

Department of Geomatics Engineering

Geomatics Engineering Programme

Thesis Advisor: Prof. Dr. Himmet KARAMAN

OCTOBER 2020

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SOSYAL MEDYA VERİLERİNİN AFET OLAYLARININ HARİTALANMASI İÇİN DEĞERLENDİRME MODELİ

DOKTORA TEZİ

Ayşe Giz GÜLNERMAN GENGEÇ
(501142609)

Geomatik Mühendisliği Anabilim Dalı

Geomatik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Himmet KARAMAN

EKİM 2020

Ayse Giz Gulnerman Gengec, a Ph.D. student of İTU Graduate School of Science Engineering and Technology student ID 501142609, successfully defended the thesis entitled "SOCIAL MEDIA DATA VALUATION MODEL FOR DISASTER INCIDENCE MAPPING", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Himmet KARAMAN** İstanbul Technical University	.....
Jury Members:	**Assoc. Prof. Dr. Turan ERDEN** Istanbul Technical University	.....
	Assoc. Prof. Dr. Bahadir ERGUN Gebze Technical University	.....
	Prof. Dr. Ergin TARI Istanbul Technical University	.....
	Prof. Dr. Stuart MARSH University of Nottingham	.....

Date of Submission : 15 September 2020
Date of Defense : 16 October 2020

To my family,

FOREWORD

This Ph.D. thesis is a production of a large amount of endeavor that belongs to me and the great people that I have met during the journey. Firstly, I would like to express my deepest gratitude and thanks to my supervisor Prof. Dr. Himmet KARAMAN for his limitless help, guidance, kindness, and patience. I would like to thank him for encouraging my study, his advice always enlightened my academic path. I would also like to thank my committee members Assoc. Prof. Dr. Turan ERDEN and Assoc. Prof. Dr. Bahadir ERGUN for their constructive criticism and contributions. I would like to thank Prof. Dr. Ergin TARI and Prof. Dr. Stuart MARSH for participating in my defense committee and their valuable comments and questions.

I would like to express my sincere gratitude to Dr. Serdar BILGI for his support and trust. I would like to thank all the members of the Istanbul Technical University Geomatics Engineering Department where I work as a Research Assistant since 2014.

I would like to thank Prof. Dr. Stuart MARSH who invited me to the University of Nottingham. I am grateful for his support, hospitality, and positivism. I had a great academic experience during my time at Nottingham Geospatial Institute (NGI) and had the chance to meet great people thanks to his working group. I would also like to thank Dr. Anahid Basiri from UCL-CASA for her immense support during my study abroad. I learned from her a lot to keep going on my research and also academic life. I would like to thank research fellows; Dr. Jessica WARDLAW and Dr. Zoe GARDNER for their kind support and friendships during my research in NGI.

I would like to thank my dear friend Dr. Elif Can CENGIZ for encouraging me to start my academic career by even collecting my official application papers. I would like to thank my lovely friends; Dr. Adalet DERVISOGLU and Dr. Fulya Basak SARIYILMAZ for always being there to share, listen and support. I am grateful to have my amazing friends Mehmet ISILER and Busra KARTAL for their friendship and sense of humor that cheer me up always. I would like to thank all my lovely friends at Nottingham University, Jialin XIAO, Loulia PEPPA, Chenyu XUE, Brian WEAVER, Kai GUO, Vivek AGARWAL, Zhengyuan QIN, Mohammed HABBOUB, Muhammad AMMAR, Direnc PEKASLAN, Damla KILIC and Tuba NAYIR for being parts of my life in Nottingham with their friendship.

Last but not least I would like to thank my great family, my mother Zeliha GULNERMAN, my father Mustafa GULNERMAN, my brother Mehmet Dogac GULNERMAN for always believing in me. I would like to thank my mother-in-law Ayse Ferah GENGEC, my father-in-law Ahmet GENGEC, my sister-in-law Sumeyye KADER for being my big family with their full support. I would like to thank my soulmate Necip Enes GENGEC for being always there with his endless support.

OCTOBER 2020

Ayse Giz GULNERMAN GENGEC
(Urban Planner)

TABLE OF CONTENTS

Page

ABBREVIATIONS

API : Application Programming Interface
CRED : Centre for Research on the Epidemiology of Disasters
DM : Disaster Management
DYFI : Did You Feel It
EM-DAT : Emergency Events Database
FEMA : Federal Emergency Management Agency
HAZTURK : Hazards Turkey
ML : Machine Learning
NB : Naïve Bayes
NCGIA : National Centre for Geographic Information and Analysis
NLP : Natural Language Processing
NNET : Neural Network
PPGIS : Public Participation Geographic Information System
SM : Social Media
SMD : Social Media Data
SQL : Structured Query Language
SVM : Support Vector Machine
VGI : Volunteered Geographic Information

LIST OF TABLES

Page

LIST OF FIGURES

Page

SOCIAL MEDIA DATA VALUATION MODEL FOR DISASTER INCIDENCE MAPPING

SUMMARY

Social Media is a new age of data sources that emerged in the last decade. Users who have diverse different motivations (such as; entertainment, communicating or promoting) sign up the platforms worldwide. Currently, there is 3.5 billion active social media account worldwide. This growing number of account holders are accepted as human sensors that provide information about their environment. Unlike the traditional sensors, these human sensors have no certainty in their capacity to sense and share the information. In addition, the data provided by human sensors is unstructured. Still, social media is an invaluable data source for studies, especially that require continuous and real-time data widely. Currently, the data is widely used for politics, marketing, and most importantly in crisis management. In this thesis, social media data is assessed for incidence mapping during or shortly after a disaster with the motivation of increasing resilience to the expected major earthquake in Istanbul.

The disaster management cycle has four phases as response, recovery, mitigation, and preparation. In the response phase, having real-time data from the affected area is important to properly allocate the resources. The conventional mapping technologies such as remote sensing and photogrammetry have the capacity of detecting the occurrence of a natural hazard however they are not eligible for information retrieval about the impacts of the natural hazards on human life such as emotions, opinions, and emergency situations. At this point, social media become forward as an immediate data source for incidence mapping during the response time of a disaster. Incidence mapping for resources management requires fine-grained data analyses. However, the uncertainty in data capacity, questions in the reliability of chosen techniques for pre-processing, and data bias are the key obstacles to the fine-grained analyses with the use of social media data. In this thesis, social media data is evaluated in terms of these key obstacles for Istanbul City since the data varies to the area that belongs to depending on its own human sensors.

The main objective of this thesis is the determination of social media data potential for its use during the response phase of disaster management. There are three sub-objectives in order to reach the main objective; revealing the adequacy of the data for incidence mapping, adapting the pre-processing steps to Turkish language and questioning the reliability of the used filtering and classifying techniques with the quantification of its impacts on mapping, and investigating the intrinsic quality of the data (such as anomalies, trends, and biases) for the further interpretation of the incidence maps.

The thesis is composed of three papers tackling these three objectives. Istanbul City is determined as the case area of each paper. In the first paper, the capacity of social media data to detect incidences in a fine-grained spatiotemporal perspective is investigated. For the case, the coup attempt data georeferenced within Istanbul city boundary is used and a series of incidences by the hour is mapped with the hotspots.

According to that study, it is revealed that social media data has the capacity to identify an incidence with a fine-grain spatiotemporal resolution. In the second paper, the reliability of the chosen techniques for pre-processing and filtering social media data is researched with its effects on incidence mapping. Two terror attacks data that is georeferenced within Istanbul City is used for the case of this study. The study is not also testing the adaptation of the current pre-processing and filtering techniques to the Turkish language and also proposes a quantitative comparative index for quantifying the spatial reliability of each filtering process. This index named Giz Index which can be replicated for the similarity searches between two incidence maps. It is found in this study, with the proposed methodology for pre-processing and filtering, over 80% of spatial reliability can be achieved for incidence mapping based on social media data. In the third paper, the intrinsic quality of data is researched for the right interpretation of the incidence maps. The study overviews the weekly sampled social media data from each month during a year that is georeferenced within the Istanbul City. The data is assessed from the perspective of data anomaly, trend, and bias with the spatial statistical tests. The study infers that the data has spatial representation bias, anomaly tendency in some parts of the city, the spatiotemporal bias in terms of the time of day and day of the week. The results of the study contribute to the incidence mapping with the reference maps to avoid biased hot spot occurrences or missing information due to less amount of data.

SOSYAL MEDYA VERİLERİNİN AFET OLAYLARININ HARİTALANMASI İÇİN DEĞERLENDİRME MODELİ

ÖZET

Web teknolojilerinin ve küresel seyrüsefer sistemlerinin gelişmesi ile ölçme, fotoğrametri, uzaktan algılama gibi geleneksel haritalama yöntemlerine bir yenisi eklenmiştir. Bu yeni yöntem; yeni coğrafya ("Neogeography"), ürün; gönüllü coğrafi bilgi ("Volunteered Geographic Information"), üreten; yeni coğrafyacılar ("Neogeographers"), gönüllüler ("Volunteers") ya da insan sensörler ("Human Sensors") olarak adlandırılmaktadır. Gönüllüler tarafından üretilen bu harita; halk katılımlı coğrafi bilgi sistemleleri, salt harita üretimi için çevrimiçi platformları ve sosyal medya platformları aracılığı ile üç farklı motivasyon ve kapsamla üretilmektedir.

Halk katılımlı coğrafi bilgi sistemleri (PPGIS), ilk defa 1996 yılında Amerika Ulusal Coğrafi Bilgi ve Analizi Merkezi (NCGIA) toplantısında literatüre geçmiştir. Kent planlamasında iki temel yaklaşım bulunmaktadır. Bunlardan ilkinde, tüm planlama kararlarını plancılar alır ve böylece kararlarda plancılar baskın olur. İkinci yaklaşımda ise, kent planlama kararları alınmadan önce halkın katılımının sağlandığı toplantılarla, halka danışılır ve planlama için tavsiyeleri alınır. Bu tavsiyelerin coğrafi bilgi sistemleri aracılığı ile toplandığı sistemlere halk katılımlı coğrafi bilgi sistemleri denilmiştir. Önceleri kağıt haritalardan halkın kentle ilgili ifade ettiği problemler ve tavsiyeler, bu sayede sayısal haritalar üzerinde ifade edilmeye ve kayıt edilmeye başlamıştır. Web teknolojisinin de gelişmesi ile günümüzde bu tip projeler çevrimiçi platformlar aracılığı ile sunulmakta ve halkın tavsiye, şikayet ve görüşleri alınabilmektedir. Çevrimiçi halk katılımlı haritalama projeleri scistarter, zooniverse ve ushahidi gibi platformlar aracılığı ile tanımlanabilmekte ve yurttaş bilimi olarak da isimlendirilen süreli projeler bu platformlar üzerinden gerçekleştirilebilmektedir.

Gönüllü coğrafi verinin ikinci türünde ise halihazır harita üretimi yapılmaktadır. Salt harita üretimine katkıda bulunmak isteyen gönüllüler, Open Street Map, Google Map ya da Here Map gibi çevrimiçi harita sunan platformlardaki eksik ya da güncel olmayan haritaları tamamlamak için yine bu platformlarda sunulan harita üretim ve sunum araçlarını kullanırlar.

Sosyal medya, gönüllü coğrafi verinin üretildiği ve sunulduğu en son çevrimiçi platform türüdür. 2020 yılı itibari ile dünyada yaklaşık 3,5 milyar sosyal medya kullanıcısı bulunmaktadır. İnsan sensörü olarak adlandırılan bu kullanıcılar, dünyanın dört bir yanından çeşitli birçok konuda veri sağlamaktadır. Sosyal medya platfomları bu kullanıcı kapasitesi ile gerçek zamanlı, sürekli ve geniş kapsamlı veri sağlama kabiliyetine sahip olmuştur. Halihazırda, sosyal medya verileri, politika, pazarlama ve en önemlisi afet sonrası vaka haritalaması gibi hayati önemi olan konularda kullanılmaktadır.

Yer bilimcilerin çalışmaları, İstanbul ilinde büyük bir deprem olacağını işaret etmektedir. Çok sayıda bina yıkım ve hasarına neden olacağı tahmin edilen bu deprem,

insan hayatını da etkileyen birçok acil durum vakasını ortaya çıkaracaktır. 18 milyona yakın insanın yaşadığı İstanbul'da bu deprem nedeniyle oluşan acil durumları hızlı bir şekilde gerçek zamanlı olarak haritalayabilecek bir sistem yoktur. Küresel anlamda mevcut olan acil durum çağrı merkezlerinin haritalama ile entegrasyonu ülkemizde de sağlanmış olsa da telekomünikasyon alanında çağrı alma, kabul etme kapasitesi sınırlıdır. Bunun yanında bu çağrıya yanıt verebilme kapasitesi personel kapasitesi ile sınırlıdır. Vaka haritalamasındaki bu gibi yetersizlikler göz önünde bulundurularak, bu tez çalışmasında, sosyal medya verilerinin afet sonrası vaka haritalaması üzerine kullanımı araştırılmıştır.

Afet yönetimi döngüsünün, müdahale, iyileşme, zarar azaltma ve hazırlık olmak üzere dört aşaması vardır. Müdahale aşamasında, afet alanı ile ilgili gerçek zamanlı verilere sahip olmak, müdahale kaynakların doğru bir şekilde tahsis edilmesi için önemlidir. Uzaktan algılama ve fotogrametri gibi geleneksel haritalama teknolojileri, doğal afetleri tespit etme kapasitesine sahiptir. Ancak, doğal afetlerin insan yaşamı üzerindeki etkileri (duygu, görüş ve acil durumlar vb.) hakkında bilgi almak için uygun değildir. Sosyal medya, bu noktada bir felaketin müdahale süresi boyunca vaka haritalaması için bir veri kaynağı olarak öne çıkmaktadır. Kaynak yönetimi için vaka haritalaması, ayrıntılı veri analizleri gerektirir. Bununla birlikte, veri kapasitesindeki belirsizlik, ön işleme için seçilen tekniklerin güvenilirliğiyle ilgili sorular ve veri önyargısı, sosyal medya verilerinin kullanımı ile hassas analizlerin önündeki en önemli engellerdir. Bu tezde, İstanbul için üretilen sosyal medya verileri, bu önemli engeller açısından değerlendirilmiştir. İstanbul için yapılan bu özel değerlendirme, sosyal medya verilerinin üretiminde yer alan insan sensörünün bölgeye göre değişmesidir.

Bu tezin temel amacı, afet yönetiminin müdahale aşamasında kullanımı için sosyal medya veri potansiyelinin belirlenmesidir. Ana amaca ulaşmak için üç alt hedef belirlenmiştir. Bunlardan ilki, insidans haritalaması için verilerin yeterliliğini ortaya koymaktır. İkinci alt hedef, ön işleme adımlarını Türkçe'ye uyarlamak ve kullanılan filtreleme ile sınıflandırma tekniklerinin haritalama üzerindeki etkilerinin nicelleştirilmesiyle güvenilirliğini sorgulamaktır. Üçüncü hedef ise sosyal medya veri kalitesinin incelenmesidir. Bu kapsamda, sosyal medya verileri ile izlecek bölge için verideki anomaliler, eğilimler ve taraflılıklar belirlenerek, vaka haritalarının daha doğru yorumlanması amaçlanmaktadır.

Bu tez çalışması, tezin kapsamındaki her bir amacı elen alan üç makaleden oluşmaktadır. Her makale için İstanbul ili vaka alanı olarak belirlenmiştir. İlk makalede, sosyal medya verileri ile ince taneli (yüksek zaman-mekansal çözünürlüklü) vaka tespit etme kapasitesi incelenmiştir. Bu durumda, İstanbul şehir sınırları içinde coğrafi referanslı darbe girişimi verileri kullanılmış ve sıcak noktalarla saat başına bir dizi olay haritalanmıştır. Bu çalışma, sosyal medya verilerinin, vaka haritalanmasında ince taneli mekansal belirleme kapasitesine sahip olduğunu ortaya çıkmıştır. İkinci makalede, sosyal medya verilerinin ön işlemesi ve filtrelenmesi için seçilen tekniklerin güvenilirliği, vaka haritalaması üzerindeki etkileri ile araştırılmıştır. Bu çalışma için İstanbul'da meydana gelen iki terör saldırısının verileri kullanılmıştır. Çalışma, mevcut ön işleme ve filtreleme tekniklerinin Türk diline uyarlanmasını test etmekte ve ayrıca her bir filtreleme işleminin mekansal güvenilirliğe etkisini ölçmek için nicel bir karşılaştırma endeksi önermektedir. Giz indeksi adı verilen bu mekansal benzerlik indeksi, iki vaka haritası arasındaki benzerlik arayışları için farklı konularda da kullanılabilir. Bu çalışmada, ön işleme ve filtreleme için önerilen yöntemle, sosyal medya verilerine dayalı vaka haritalaması için %80 üzerinde mekansal güvenilirliğe ulaşılmıştır. Üçüncü makalede, İstanbul mekansal alanı için paylaşan sosyal medya

verileri incelenmiştir. Bu araştırma ile vaka haritalarının doğru yorumlanması için referans haritalarının üretimi hedeflenmiştir. Çalışma, yılın her bir ayına ait bir haftalık örneklenmiş sosyal medya verilerini gözden geçirmektedir. Örneklenmiş bu veri, uzamsal istatistiksel testlerle veri anomalisi, eğilim ve yanlılık açısından değerlendirilir. Çalışma, verilerin mekânsal temsil yanlılığı, kentin bazı bölgelerinde anomali eğilimi, günün bazı zaman dilimleri ve haftanın bazı günlerine göre mekânsal-zaman yanlılığına sahip olduğunu göstermektedir. Çalışmanın sonuçları, veri yanlılığından kaynaklı olarak sıcak nokta oluşumlarında ve/ veya daha az miktarda veri nedeniyle göz ardı edilme olasılıklarını önlemek için referans haritalarıyla vaka haritalamasına katkıda bulunmaktadır.

1. INTRODUCTION

Social media (SM) is the advent of web 2.0 technology and the platforms have been ubiquitously escalated in the last decade (Goodchild, 2007a; Hall et al, 2010). It has developed with the features adding such as; sharing media, conference calling and rating the shared posts (Moffitt, 2017). The features of georeferencing and geotagging are the two important embeddings of the SM which mark an era for mapping widely by the help of human sensors (Zhao et al, 2011). Technological developments in SM platforms and the widespread of smart mobile devices and the internet let a growing number of users in these platforms. Nowadays, SM has over 3.8 billion users who are forming a continuous, wide, real-time data source worldwide (Statista, 2020).

The use of social media platforms varies according to the users' motivation as the platforms are designated in terms of different purposes such as; news tracking as on Twitter, social networking with acquaintances as on Facebook, and photo sharing as on Instagram (Li et al, 2013; Poell & Borra, 2011). These purposes are supported by the social networking design of the platforms and the features of content sharing.

Social media data (SMD) is a crowdsourced data that include diverse topics such as; politics, celebrations or daily things, etc. (Issa et al, 2017; Kim et al, 2016; Lansley and Longley, 2016; Li et al, 2013). While this let conducting a wide variety of studies, at the same time this makes each study hard due to its unstructured and unclassified data content apart from the traditional data sources. Yet, SM become an invaluable data source since its continuous and wide data producing capability.

SM enabled new age of mapping as a new and fostered part of neogeography which is sourced by neogeographers (Turner, 2006). The terms lately called volunteered geography sourced by the volunteers (Elwood et al, 2012; Goodchild, 2007a, 2007b, 2009). These new terms embody the map production by volunteers who has no expertise in mapping with the tool geographic information systems. Although the forms of the platforms and the aim of the projects vary and lead changing terminology over volunteered geographic information (VGI), the last product (map) and the source (volunteers) are not differentiated at the end.

SM platforms are the last subdivision of the VGI sources after public participation platforms and peer map production platforms (Abbasi & Liu, 2013). Data providers are the users who allow their activities to be seen publicly with the signed terms and conduct while they are using the platforms either consciously or unconciously. Hecht and Shekhar (2014) called SM users as unconscious volunteers in the context of VGI and these unconscious volunteers who enabled their location-based services while posting on SM become a part of a mapping study without any awareness of the projects.

SM-VGI has several advantages and disadvantages apart from other VGI types. First, it serves data without any need for project deployment of a specific topic while this might mean an uncertain amount of related data at the same time. In respect to this, there might be either abundant data or almost none. Second, SM-VGI includes various content that can help to analyze different aspects of a study, on the other hand, this unclassified data should be classified to determine relevancy to the topic that is researched. The last advantage of the SM-VGI is continuity of data production worldwide thanks to its non-stop server and a massive number of human sensors already signed in to the SM platforms. However, data produced have several biases such as; population, representation, temporal and spatiotemporal, etc.

The most important part of the success of SM-VGI is game theory as a way of attracting more human-sensors into the SM platforms. In this way, SM could become a major data source for human-based research with its no-cost growing number of sensors worldwide, unlike other VGI projects. Former VGI types have limited participants for determining topics, time intervals, and spatial boundaries that serve structured data for pre-designed specific purposes. As a result, while former VGI studies and platforms maintain data production for a specific manner, SM-VGI would provide data even it is not presumed before. For this reason, SM-VGI is an important source for many human-based projects such as politics, marketing, and urban planning but more importantly, is vital for disaster management in every kind that affects the environment and human life.

SMD has two main components as text and location (latitude, longitude). There are tremendous studies on SMD with the expectation of information retrieval on different topics from different aspects. Initial studies are mostly on text component since the features of georeferencing and geotagging added lately to SM platforms. Those studies

are to determine the trend topics which is an easy task when SM reacts to a big event such as; a festival, a political event or a catastrophic disaster. However, information inferring while there are minor incidences such as secondary incidences depending on a major one, might be a big issue due to unstructured content. In respect to this, SMD requires filtering and classification steps in other words structuring the content, for further processing. Without these steps, SMD can not be known whether it is relevant to the topic (domain). The relevancy of the content is determined by the techniques that is chosen for filtering or classification. Therefore, the accuracy of the relevancy determination depends on the reliability of the chosen structuring techniques. Besides that, credibility is another aspect of the SMD which is mostly assumed manipulated easily. This means that data includes misinformation due to the sensor (a human, a bot or a company team), therefore it is mostly measured and named as credibility of data providers.

In the context of SM-VGI, the text component of the data can be assumed as the attribute part of each spatial object (i.e. point). Beside the data quality issues of text component, the location component of the data is another aspect that should be assessed furtherly. At first, SMD location component can be attached in several different ways in terms of the volunteers' motivation or consciousness that might mislead the analysis. As in text mining for determining trend topic, hotspot mapping might not be a big issue while there is abundance of data as a reaction to a big event. However, when a fine-grained spatiotemporal analysis is aimed in order to detect minor events, the additional data mining processes are required such as crosscheck validation with the other data providers or credibility control of users. In addition, while mapping with the use of SMD, filtering process has impacts on data. In other words, the reliability of filtering techniques affects the accuracy of retrieved spatial information. Also, SMD might include several types of bias (such as; representation, temporal, and spatial) since sensors have no common standards on data production as in traditional sensor systems for mapping.

In this thesis, SMD is assessed in the perspective of incidence mapping regarding data capability, reliability and bias to meet the vital role of SM during and short after a major disaster.

1.1 Research Motivation and Problem Space

In the last two decades, communication technology and parallel to data acquisition techniques for mapping have been evolved in the last half-century. One of the developing mapping technologies is remote sensing satellites with their increasing resolution (spatial, temporal, radiometric) capability day by day. Yet the acquired data either via satellites images have provided capability to understand the occurrence of the natural hazard, they are not capable of retrieving the information of natural hazards impacts on human life including emotions, opinions and emergency situations (wounded, death or stacked people due to a catastrophic manmade or natural disasters).

Telecommunication is another data source of emergency situation mapping, and some countries have distributed call centers for each emergency services (such as; fire department, ambulance, and police) while some others have centralized the call centers for all types. Either the emergency call services are distributed or centralized, having all calls within short time after a catastrophic disaster is not possible in reality. These centers require huge staff capacity and high organization capability in short time to accept the call. In addition, telecommunication infrastructure should be capable of carrying the workload of immense calls to call centers and also between people for checking their families, relatives, friends. After a catastrophic disaster affecting large areas and covering the most populated regions of the world, current telecommunication services suffer unable to help respond.

There are local attempts to build a communication infrastructure for disaster management during or shortly after a disaster via public participation. Life-Saving Kiosk project is one of them and proposes to collect data from the public with the allocated digital smart kiosk across to the accessible points of neighborhoods (can be pointed as the gathering area) within a city. This project is a very organized way of data collection while there is a catastrophic event. On the other hand, building such a system and maintenaning it can be very expensive.

Crowdsourcing for humanitarian circumstances is another smart way of the emergency mapping project. Humanitarian OpenStreetMap Team (HOT) is an international non-profit organization that carries out data digitizing projects over DigitalGlobe satellite images to provide base maps where it is needed shortly after a disaster. HOT contribute to the production of missing maps and also to building a local community for easying

disaster relief as in Nepal, Katmandu Living Lab Project. However, this organizing a team for map production and gathering a local team takes time. Although the contribution of completing the missing maps is enormous for undeveloped regions, this organization is not directly aiming the incidence mapping during or shortly after a disaster for developed and highly populated regions.

SM is the last developed communication technology and has been used in disaster management with diverse tasks for each disaster management phase (preparedness, mitigation, response and recovery) in the last decade (Houston et al, 2015). Unlike previous technologies used for data acquisition, SM has been naturally there collecting and providing data in any circumstances uninterruptedly (if there is no censorship and big system failure) with the human-based sensors. Therefore, SM has pioneering potential on catastrophic disasters management as a real-time, continuous, world-wide, fast data provider without extra cost and organization plan for building up a data collection team.

The value of SMD is much more significant to handle major disasters that occurred in big cities where millions of people can be affected in many ways. Istanbul is one of the most populated cities in the world and highly seismically active area. According to (Bilham, 2010; Parsons, 2004; Parsons et al, 2000), Istanbul is expecting a major earthquake that has more than 7.0 magnitude within the following 30 years. Impacts of this earthquake are simulated across Istanbul via hazard estimation software with regards to the building age, height and structure, and the ground micro-zoning maps in the city. According to those probabilistic hazard or loss estimations, large amount of buildings will be demolished or moderately damaged (Erdik et al, 2008, 2011; Konukcu et al, 2016). This will obviously cause a chaotic situation at the first moment. Following that, the race against time will start in order to save lives by coping with secondary incidences stimulated by the main disaster such as fire, flood, and debris.

Karaman and Erden pointed real time earthquake hazard maps require complex calculation algorithms including geospatial information related to the region of interest and longer time. Rapidity is the most important part of the response and recovery phases of the disaster management cycle. Losing time to acquire accurate disaster management information for these phases is intolerable (Karaman & Erden, 2014). Therefore, a plan to allocate emergency services should be designed in the preparation phase of the disaster management cycle. In addition to that, the reality should be

considered while the affected fields require emergency services. In respect to that, a well-designed plan could only be implemented if there is real-time information from the affected field in the response phase of the disaster management cycle. Consequently, the problem is the requirement of real-time data from the affected area where there is a catastrophic disaster. Unfortunately, the technologies that are previously used for information retrieval were not adequate to meet the requirements of human-related incidences as mentioned above. For this reason, SMD is assessed as a data source for incidence mapping in this thesis.

This thesis is designed as a composition of three academic papers. The first one is including comprehensive literature review on use of social media data for disaster management and a case study to show information inference capability to use in general concept. The second one is proposing a text mining approach to filter out irrelevant contents with sentimental analysis over Turkish texts. The third one is tackling the intrinsic data quality of SMD for a determined case area and model a data investigation methodology in order to increase the right interpretation of incidence mapping in a fine-grained spatial perspective.

Mostly Loss Assessment Studies, are required some inputs data like age of building, building structure and height of building and these kinds of studies accuracy depends on the reliability of inputs. In addition, those kinds of studies have mainly reflected the situation to manage disaster in the period of mitigation and preparedness, and to figure out the general view of earthquake impacts within response and recover period. On the other hand, reality of events is required to intervene for response and recovery period of disaster management. In this study, Loss Assessment and road network functionality will be modelled through Social Media data for earthquake response and recovery period management.

This study is important to avoid negative and chaotic vital, social and economic consequences of expected major Istanbul Earthquake. Furthermore, the methodology and algorithms of the proposed study is expected to contribute to Geographic Information System and Data Science, Data Mining and Spatial Statistics Applications.

1.2 Research Question

SM appears to be a pioneering option to provide data for incidence mapping that is vital shortly after a disaster. However, SMD should be assessed in terms of its capacity, reliability and intrinsic data quality for its use in this vital mission. In respect to this, research questions for this thesis are determined as below.

1- "Does SMD has the real capacity to extract information for incidence mapping?"

2- "Are the techniques used for filtering and classifying SMD reliable and how these techniques impact the spatial reliability of incidence mapping?"

3- "What are the anomalies, trends, and bias in SMD? How this intrinsic data quality perspective can be assessed for incidence mapping?"

1.3 Research Aim and Objectives

The main aim of this Ph.D. research is the valuation of SMD with regards to information mining methodologies in order to increase the resilience of the Istanbul Metropolitan Area against the expected major earthquake. There are three objectives that are aimed to be reached related with that. The first objective is to reveal how the spatial feature of SMD works and whether its potential might be adequate for incidence mapping. The second objective is to adapt the preprocessing steps (text cleaning and filtering) to the Turkish language and quantify the reliability of the methodology with the impacts on incidence mapping. The third objective is the investigation of data anomalies, trends, and biases in data and producing reference maps for the right interpretation of spatial anomalies when there is a disaster in the city.

The objectives of this study are determined to assess SMD in different aspects to evaluate data potential for the benefit of information extraction during the response phase of disaster management.

1.4 Research Contribution and Publications

SMD is used for information extraction in diverse topics especially that requires wide-area research. Studies on SMD result in a map that is generally presented as coarse-grained as country, county or city level. The performance of the techniques used for

extracting this spatial form of information is quantified with the administrative information. This kind of quantification might return to high performance for coarse-grained analyses with the abundance of data. However, SMD requires much attention than this to process it for fine-grained spatial analyses since the results might be misleaded due to its intrinsic and extrinsic data quality and reliability of chosen methodologies.

This study contributes to the valuation of SMD mining for fine-grained incidence mapping with the three published papers. Each publication tackles the different aspects of the data and processing methodologies for fine-grained analyses.

The first publication presents a review of SMD use for DM and a case study to show the capacity of SMD for incidence mapping. The case study reveals how spatial data can be manipulated by the account holder and the spatiotemporal monitoring capacity of incidences during a man-made disaster.

The second publication tackles reliability of SMD filtering technologies and their impacts on incidence mapping. The study focuses on the two main investigations; 1- the use of common approaches for domain-based filtering Social Media Data (SMD) in the Turkish Language, 2- the spatial reliability of the incidence maps which are produced with the domain-based filtered SMD. The study is important and novel since it presenting the non-English language filtering details with the exploratory analyses and proposing the mapping reliability measurements with a new similarity index entitled Giz Index, which is designed specifically for incidence maps apart from the current similarity indexes. In addition to this, the study points out the reliability discussion on and reliability quantification of maps, which is the last outcome of many social media studies. In this way, this study contributes to the advances in SMD filtering and assessment of incidence mapping analysis. Another novelty of this study is to filter and validate the SMD in a rapid succession to evaluate the social media data spatially for emergency cases.

The third paper searches on the intrinsic data quality of SMD within a spatial boundary and proposes a methodology for data investigation. This investigation has three main following steps; spatial anomaly detection, production of trend map, and spatiotemporal bias assessment. The methodology contributes to fine-grained metropolitan area monitoring systems with its normalization techniques on

representation bias, and proposals on weights of spatiotemporal bias assessment. The study is novel since the methodology normalizes the data that comes from the highly represented users instead of removing them. In this way, reference maps have normalized noise and will help to degrade the noise in further analyses.

1.5 Research Theme and Structure

Themes of this thesis organised as based on the three publications that are a book chapter and two SCI- Expanded articles as follows:

Chapter 1 consists of introduction, research motivation, research question, research aim, and research contribution and publication subsections.

Chapter 2 presents the book chapter entitled “New Age of Crisis Management with Social Media”.

Chapter 3 presents the journal article entitled “Spatial Reliability Assessment of Social Media Mining Techniques with regard to Disaster Domain-Based Filtering”.

Chapter 4 presents the journal article entitled “Citizens’ Spatial Footprint on Twitter: Anomaly, Trend and Bias Investigation in Istanbul”

Chapter 5 presents the conclusion of this study and recomendations for further studies.

2. NEW AGE OF CRISIS MANAGEMENT WITH SOCIAL MEDIA[1]

2.1 Abstract

Social Media (SM) Volunteered Geographic Information (VGI) is gradually being used for representing the real-time situation during emergency. This chapter presents the SM-VGI review as a new age contribution to emergency management. The study analyses a series of emergencies during the so-called coup attempt within the boundary of Istanbul on the 15th of July 2016 in terms of spatial clusters in time and textual frequencies within 24 hours. The aim of the study is to gain an understanding of the usefulness of geo-referenced Social Media Data (SMD) in monitoring emergencies. Inferences exhibit that SM-VGI can rapidly provide the information in the spatiotemporal context with the proper validations, in this way it has advantages to use during emergencies. In addition, even though geo-referenced data embody the small percent of the total volume of the SMD, it would specify reliable spatial clusters for the events, monitoring with optimized-hot-spot analysis and with the word frequencies of its attributes.

Keywords: Social Media, Volunteered Geographic Information, Disaster Management, Spatial Data Mining, Text Mining

Acknowledgments This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK-2214/A Grant Program) and Istanbul Technical University Scientific Research Projects Funding Program (ITUBAP-40569).

[1]This chapter is based on a book chapter: Gulnerman, A. G., Karaman, H., Basiri, A. (2020), New age of crisis management with social media, *Open Source Geospatial Science for Urban Studies* - Lecture Notes in Intelligent Transportation and Infrastructure, Springer, ISBN: 978-3-030-58232-6. doi: https://doi.org/10.1007/978-3-030-58232-6_8 .

2.2 Introduction

A disaster is defined as an emergency condition, which turns into serious casualties when it exceeds the capacity of available resources to manage it. All activities within this management are to avoid the loss of life and money (FEMA, 2000) International Federation of Red Cross and Red Crescent Societies declare a disaster is "a sudden, calamitous event that seriously disrupts the functioning of a community or society and causes human, material, and economic or environmental losses that exceed the community's or society's ability to cope using its own resources" (IFRC). The Emergency Event Database (EM-DAT) which was launched by the Centre for Research on the Epidemiology of Disasters (CRED) and initially supported by the World Health Organization (WHO) and the Belgian Government (CRED, 2016). That global disaster event database indexed disasters by conforming at least one the following criteria; 10 or more people dead, 100 or more people affected, the declaration of a state of emergency, call for international assistance (EM-DAT, 2009).

(Alexander, 1993) classified the natural disasters with respect to many aspects like duration of impact, length of forecasting, frequency or type of occurrence etc. EM-DAT classifies disaster under two main branches as "natural" and "technological" (Abbasi & Liu, 2013). While EM-DAT has no subcategory for a terrorist attack as disaster or emergency, (Berren et al, 1980) categorises them in a five-dimensional approach as type of disaster, duration of disaster, degree of personal impact, potential for occurrence and control over future impact. They focused on the type of disaster part on the non-natural disasters as manmade events like holocaust, kidnapping, and plane crashes that have affected mass amount of people like a coup event. (Johnson, 2000) categorizes terrorism under the technological categories in his whitepaper. Disasters arise from hazards, which can be classified into three main categories of origin: natural, technological and environmental degradation. While natural disasters can originate from hydrometeorological, geological and biological hazards, environmental degradation is often induced by uncontrolled and unplanned human activity in nature.

This chapter focuses on a branch of technological hazard that originated from "*internal disturbances planned by a group or individual to intentionally cause disruption like*

riots, violent strikes, and attacks, including the act of large-scale terrorism" as (Johnson, 2000) mentions in his whitepaper. (Houston et al, 2015) also used a likely

classification for the disasters as (Johnson, 2000) does by their causes like; natural (such as an earthquake or a hurricane), technological (such as an oil spill), or human (such as terrorism). He also takes into account the consequences of the disasters while mentioning the political consequences too. (Eshghi & Larson, 2008) have also taken into account the Canadian Disaster Database classification with five classes as biological such as an epidemic, geological, such as an earthquake, meteorological and hydrological such as a drought, human conflict such as terrorism and technological such as chemical materials. They also referred the Disaster Database Project by (Green, 2003) where he classified the disasters as; conflict-based disasters like bombing and massacre, human system failure like dam collapse and mine accident and natural disasters like earthquakes, storms etc. World Health Organization's Emergency Health Training Program for Africa classifies the disaster hazards as Natural – Physical as external like topographical or internal like tectonics and telluric, Natural – Biological as epidemics or infestations and Manmade / Technological as industrial, nuclear, chemical, fires, wars or civil strife and structural failures (WHO & EHA, 1999). In addition to that, Terrorism as a disaster and an emergency issue have been assessed by (Cutter, 2003) from the perspective of Geographical Information Science. The study draws perspective from both emergency management practitioners and first responders (such as; police, fire emergency, and medical teams) and emphasize the critical need of real-time data from the field.

All phases of Disaster Management (DM) that are preparedness, mitigation, response and recovery are crucial. However, the response phase can be seen the most crucial one due to race against time for the use of scarce resources. While all levels of the authorities have responsibilities to facilitate the management periods, communities, non-governmental organizations and individual contributors play important roles in providing information pertain to disaster effects on the field for this phase (FEMA, 2000, 2007; Gulnerman et al, 2017; Karaman & Erden, 2014).

The first requirement of Emergency Management is information about the affected area. Most of the disaster management systems and programs count on the estimations that are compiled before the hazard event occurred at the location like HAZUS (Schneider and Schauer, 2006), HAZTURK (Elnashai et al, 2008; Karaman et al,

2008) and ELER (Hancilar et al, 2010). Most of the estimations encounter uncertainty and miscalculation because of the deterministic or probabilistic scenario creation at the start (Elnashai et al, 2008; Erdik et al, 2008; Karaman et al, 2008; Kircher et al, 2006). The rest of the disaster management systems count on visual interpretation, digital vectorization parts following the disaster and this takes more time than the DM can spare since a rapid response is the most important part of the DM, especially in the response and recovery phases (Karaman and Erden, 2014). In addition to that, (Goodchild, 2007a) questioning the capability of remote sensing for the detection of the emergency situations by asking "Could every emergency be sensed by satellites?". Beside the statistical and remote sensing techniques for disaster crisis management, the new era has started with the internet technology. One of the very first examples of the citizen science project on earthquake hazards is "Did You Feel It?" (DYFI) which is introduced in 1999 (USGS, 2019). It is a pioneering automated method to collect macro-seismic intensity data from internet users' shaking and damage reports. The system aims rapid information collection which is highly required for emergency during earthquakes (Wald et al, 2012). However, this kind of project have limited by determined area and type of disaster.

The evaluation of technology in disaster management has started with the technological invent of Geographical Information Systems (GIS) in 1960s (Clarke, 1997), continued with Peer-Production Volunteered Projects (HOT) and Citizen Science Volunteered Projects (Hirata et al, 2015; Meier & Werby, 2011) thanks to widespread use of internet in 2000s. Further developments were brought with SM platforms in late 2000s that marked a new epoch in disaster management (Acar and Muraki, 2011; Goodchild, 2007b; Ishino et al, 2012; Issa et al, 2017; Iwanaga et al, 2011; Sakaki et al, 2010; Z. Wang et al, 2016). Presently, SM has more than 2 billion active bionic sensors (users) who sense and share what is happening around them (Statista, 2017a). They produce real-time data from most of the world where has internet infrastructures and no censorship.

Volunteered Geographic Information (VGI) is stated as crucial for rapid information gathering from disaster area regarding time scarcity and limited resources in emergencies. Initial aim of this study is to draw the perspectives of VGI and its potential contribution to the emergency management. In respect to that, introduction section is extended with three sub sections. In the first part, differences between

"volunteers", platforms used for contribution and motivations behind their willingness are discussed. Second, the study literates Social Media (SM) - VGI to demonstrate the potential use in different kind of DM projects. Third, SM platforms and their added VGI features are mentioned.

Following that, a case study is conducted with SM-VGI in order to show data accountability, availability and incidence detection capacity of the SMD over space and time. The case is related with so-called coup attempt in Turkey. On the 15th July 2016 Turkey came under attack by putschists. Official sources announced that there were 250 deaths and more than 2,000 injuries during this coup attempt. The attempt has been assessed as a series of emergency events under the title of "terrorism" within types of disasters. Since all traditional media sources were silent during that time, only social media served citizens to be informed and communicate each other. The coup attempt covered the whole country but for the government and emergency response teams, it was not easy to handle or manage this disaster. This indicates that sharing items on SM is valuable during such an event for both citizens, emergency response teams to be informed in real-time.

The case section is divided into four subsections. In the first subsection, the spatial variation of SMD is tested with several sharing specifications in order to find out changing spatial data accountability. In the second sub, plenty of SMD capturing tools are introduced in terms of their different purposes. The third subsection presents case data as SMD that is the only available data for the critical first hours of the attacks. In the last part of the case study, data has been basically analysed by spatially and textually to gain more insight about data evaluation in time during that kind series of emergency conditions. The fourth main section includes the debates for validation of the data by matching news collected from the trustable newspaper or TV channel sources. Lastly, the study concluded with the general overview of potentials and weaknesses of SM geo-referenced data and further required studies to enhance the understandings of data reliability.

2.2.1 Volunteers of VGI

The basic questions about the SM users are; "who are they?" and "how did they emerge?" The answer is brought by VGI terminology. They are the volunteers, evolved with technology, and have several different contextual names such as; participators,

volunteers, and neo-geographers (Goodchild, 2009; B. Hecht & Shekhar, 2014). Contexts of the VGI and its varying termed volunteers is structured in Table 2.1 to provide easier understanding for the literature of volunteers. (Parker, 2014; Turner, 2006) and (Goodchild, 2009) called the idea of contributing to the map production with the existing online toolsets by untrained people as "neo-geography". And the untrained producers of this new mapping idea are called "neo-geographers" means all kinds of volunteers in VGI.

Table 2.1 : VGI Terminology (Elwood et al, 2012; Gulnerman et al, 2016; Hecht and Shekhar, 2014; Parker, 2014; Turner, 2006).

		Kind of Volunteers		Platforms
Citizen Science VGI		Public Actions Participators (Residents, Tourists, Business etc.)		SoftGIS, Zooniverse, Ushahidi, Scistarters
Peer- Production VGI	Neo-geography	Deliberate Volunteers (trained or untrained contributors)	Neo-geographers	Open Street Map
SM-VGI		Unconscious Volunteers (application users)		Twitter, Facebook, Google, Instagram

Initial examples of VGI are based on participatory urban planning with the use of Geographic Information Systems that is called the Public Participation Geographic Information System (PPGIS) (Anderson, 1995; Schroeder, 1996) .Volunteers involved in these kinds of studies are mostly local people (such as; citizens and employees), and semi-local people (such as; tourists or visitors) (Gulnerman and Karaman, 2015; Sieber, 2006) and they are specifically named as participators. Initially, participators embody relatively small groups of volunteers via local meetings. Later it turned into larger groups of people with online PPGIS (Ball, 2002) and even become wider citizen science projects (Hirata et al, 2015; Meier, 2012; Meier and Werby, 2011; Muller et al, 2016; Wardlaw and Jackson.; Wardlaw et al.) with the help of online platforms such as; (Zooniverse, 2017), (Scistarters, 2017) and (Ushahidi, 2017).

Deliberate volunteers are assigned as directly as main volunteers of VGI, their focus contribution is producing base maps such as; buildings, roads, places, and parks (Elwood et al, 2012; Goodchild, 2009; OSM, 2016). This second type of untrained volunteers emerged with online base map platforms (Elwood et al, 2012; Parker, 2014; Turner, 2006). Open Street Map (OSM) is a well-known platforms of this type of VGI, that was deployed in 2004 (OSM, 2016). It has also Humanitarian Open Street Map Team (HOT) projects for the humanitarian purposes (HOT).

Hecht and Shekhar (2014) list the volunteers in SM as unconscious who are unaware of their location information use. This unconsciousness is caused by the "terms of service" acceptance of users by simply clicking the checkbox to declare, "I have read and signed this agreement". In SM-VGI, volunteers share diverse content (such as; emotions, memories and news) with no structured topic limitation for a specific project. However, volunteers of SM-VGI become together around a topic and protest authorities, companies and regulations (Haciyakupoglu and Zhang, 2015; Lim, 2012; Tufekci and Wilson, 2012; Valenzuela et al, 2012).

Tulloch (2008) denoted that the overlap between Citizen Science and the Peer Production of VGI relies on the location investigation by individuals. Though the potential difference between Peer Production and Citizen Science relates to the purpose of motivation for contribution that Citizen Science Projects are implemented to inform planning and policy decision makers, though Peer Production systems may have no purpose other than volunteers' pleasure. On the other hand, (Hall et al, 2010) indicate the disparity between these two terms in Web 2.0 geospatial applications appears more semantic than real.

Goodchild (2007a), emphasize the all-human population as sensors around the world. According to that, people provide the data that they sense without any gain. On the other hand, people do not only provide data as they are. They interpret their sensations with their inner and outer visions. Ball (2002) claims that citizen science projects may include biased results because participators tend to manipulate all data that they provide for their personal benefit. Although, motivations of VGI types are different, all volunteers' data depends local knowledge which has blurred the distinction on the non-expert amateur and expert for map making (Goodchild, 2009). The advantage of SMD may arise at this point; unconscious volunteers would not manipulate the results of the studies consciously. In this study, SM-VGI is assessed as a geo-news for crisis time.

2.2.2 SM-VGI studies for DM

SM after its emergence of and rapid sprawl of internet usage has attracted many researchers' attention on that topic. Initially most of the SM studies based on the new media researches including text mining and sentiment analysis. Later, the use of Global Navigation Satellite Systems (GNSS) allowed SM platforms to be used as VGI (Gross

and Hanna, 2010; Moffitt, 2017). An only small percent of the SMD are geo-referenced with precise latitude and longitude coordinates (Power et al, 2015). However, there are geo-parsing studies to extract location information from text content with text mining methodologies. These geo-parsing techniques can be vital for the time of emergencies. Gelernter and Mushegian (2011) used a geo-parsing technique to determine the locations of tweets pertain to an earthquake. For a high-performance geo-parsing technique, Gelernter and Wu (2012) studied 30.000 posts on the 2011 fire in Texas. Moreover, Gelernter and Balaji (2013) built a heuristic algorithm based on Named Entity Recognition (NER) in order to identify the streets and addresses, buildings, urban spaces, toponyms, place acronyms, and abbreviations.

Leetaru et al (2013) spot on that there is a crucial requirement to have a better understanding of the twitter geography due to the wide use of SMD during emergencies. This is well recognised by the United Nations (UN), as the UN has produced their first crisis map solely based on SMD (Meier, 2012). Similarly, Power et al. (2015) used 1.8 billion tweets to monitor hazards in a crisis coordination center using text mining and machine learning algorithms. Landwehr et al. (2016) have mentioned advantageous and contribution of SMD to mitigate effects of the natural disasters. Utani et al. (2011) denoted that generally, hashtags (#) are being used to classify Twitter data to utilize a high-tech response system (such as; Person Finder", "Traffic Road Information", and "Relief Supply Matching System") in Japan.

SMD is mostly used for detecting the locations of emergency events. Sakaki et al. (2010) developed a real-time system to detect earthquakes and send the notifications to the registered users using SMD. MeCab analyser (Kudou, 2013) was used to categorise sentences into sets of words in Japanese for the text analysis. Another study (Ao et al., 2014) used 480,000 posts and inferred location contextually. This means inferring the location where the post is about rather than it comes from. Language word classification tools to infer the locations in the messages for the Chinese language were also used for this study (Ao et al, 2014). These two studies adopt specifically the local language text classifiers that are important for more reliable categorization and information extraction during a disaster.

SMD is also used for early warning and rescue purposes. An ongoing study on SMD is designed to support real-time early warning and planning systems for tsunamis risk in Indonesia (Landwehr et al, 2016). The study records the historical tweeting activities

to estimate the population density change over time to plan the evacuation and response. (Acar and Muraki, 2011) have tested the crisis communication on SM platforms. The results showed that Twitter is the only platform after the Great Tohoku earthquake, which was 9.0 scale Richter earthquake that hit Japan. The research investigated victims' experiences on SM during the crisis and tried to improve their communication (Acar and Muraki, 2011). By considering the rapid increase in the volume of SMD and the ability of being alternative of the potentially destroyed telecommunication infrastructures, (Yin et al, 2012) have proposed a notification system to enhance the awareness of emergency using some textual clustering techniques applied to high-speed stream data during disasters and crises. Wang et al. (2016) studied the characteristics of wildfire over space and time using SMD to gain more insight about SMD usage in situational awareness and disaster management. The study concluded that people have relatively strong geographical awareness during wildfire hazards and are likely to communicate regarding the situational informs on wildfire hazards, response and to show their gratitude to firefighters (Wang et al, 2016).

SM is accepted a pioneering communication way during disasters, on the other hand there are debates about reliability of the data. Castillo et al. (2013) have studied the SM trend topic activities after disasters such as; tsunami alerts, missing people, road conditions. The study investigated the credibility of the information considering the information propagation and false rumour propagation using heuristic-based filters (Castillo et al, 2013). Poorazizi et al. (2015) proposed a VGI quality metric to standardize the disaster management of five headings of positional nearness, temporal nearness, semantic similarity, cross-referencing and the credibility. Another quality evaluation was conducted by Crooks et al. (2013). In this study, 125,000 posts from DYFI (Wald et al, 2012) were compared with approximately 21,000 of geo-referenced tweets and concluded that Twitter data semantic filtering is limited to few words.

SMD is also used for disaster-related mapping such as; the affected area, sentiments of victims, and transportation behaviour of victims. Rosser et al. (2017) have proposed a method for rapid mapping of the flood inundation extent using the geotagged photographs shared by SM users, also remote sensing and topographic map data. Lin and Margolin (2014) looked into the emotion sprawl after a terrorist attack by analysing the SMD. The sentiment and time series analysis are applied to 180 million

geocoded tweets over a one-month period. Fear sprawl was interrogated according to locational proximity, social connection and physical connection and associated with them. Another sentiment analyses techniques were applied to 97,000 tweets for operational crises management (Terpstra et al, 2012). SMD allows to analyse the behavior users during and after the natural disasters, as shown by Hara (2015) study, where 3,307 users with over 130,000 total tweets and approximately 23,000 geo-tagged tweets were studied to estimate the victims behaviour while returning-home modes after the Great Eastern Japan Earthquake in 2011.

In addition to all, Houston et al. (2015) searched on the typology & conceptualization of the SM usage for disaster management at a number of levels. The study lists the functions of SM regarding phases of disaster management. This study can be assessed as "document and learn what is happening in a disaster" that is an initial study for the function of "provide and receive disaster response information: identify and list ways to assist in the disaster response".

2.2.3 Twitter and VGI features

There are several types of SM platforms such as micro-blogs like Twitter, discussion forums like Reddit, digital content sharing platforms like Instagram, social gaming sites like Mobage, social networking sites like Facebook (Houston et al, 2015). As literature review broadens, SM-VGI studies mostly focus on the Twitter due to the popularity of the twitter usage. Twitter launched publicly in 2006 (Moffitt, 2017). Statista (2018) declared that it has reached 330 million monthly active users as to April 2018. There are nearly 3.3 billion internet users who are active on social networks around the world, whereas active social network penetration in Turkey as of January 2018 is 63% (Statista, 2017a, 2020). Beside its growing popularity, Twitter added geo-referencing feature into its platform in 2009. This feature is named as Tweet Geotagging API that allows extracting the geographic coordinates from the tweets. Following that, Twitter announced their "Twitter Places" in 2010 that signify a geographic area on the venue, neighbourhood, or town scale (Moffitt, 2017).

Moffitt (2017) identified that twitter has three main sources to geo-reference tweets. The first one is geographical reference in a tweet message. The second is geo-tagging tweets by the selection of the user thanks to GNSS. The third one is the account profile set as home location by the user. The first one requires geo-parsing techniques for geo-

referencing that is mostly returns limited success due to several Natural Language Processing (NLP) limitations. The second one is the most accepted source of geo-referencing a tweet. The third one is directly accepting the home location for all tweets geo-referencing of the user, which is not convenient for disaster cases. Even the most usable and sufficient source of geo-referencing seems to be the second one; there are several ways of manipulating the location of a tweet via several other location-based applications and sources. In the next section, a case study is conducted to show basic disaster event mapping capacity of SMD and data accountability for tweet geo-referencing.

2.3 Case Study

2.3.1 Data accountability

Variations of tweeting preferences affects tweets' geo-referencing and it is important to know data accountability conditions for mapping. Plenty of determined preferences is applied while tweeting from the same location to map possible spatial diversity (Table 2.2). The real "user location" in this test is marked on maps (Figure 2.2) where is the location of Istanbul Technical University Faculty of Civil Engineering, the room L-201. Totally 17 tweets are posted via Twitter, Swarm and Instagram. Swarm and Instagram have link to twitter platform for posting the same content as tweet. Test tweets are listed in Table 2.2 with the id, tweet, platform, connection type and details of tweet action.

Table 2.2 : Detail of Tweets Test List.

ID	Tweet	Platform	Connection Type	Details of Tweet Action
1	Test 1	Twitter	3G	without location
2	Test 2	Twitter	3G	with location label (Istanbul, Turkey)
3	Test 3	Twitter	3G	with location label (Sariyer, Istanbul)
4	Test 4	Twitter	3G	with location label (Resitpasa, Istanbul)
5	Test 5	Twitter	3G	with location label (Maslak, Istanbul)
6	Test 6	Twitter	3G	with location label (Emirgan, Istanbul)
7	Test 7	Twitter	3G	with detailed location corresponding GPS
8	Test 8	Swarm	3G	Insaat fakultesi selection
9	Test 9	Swarm	3G	ITU selection
10	Test 10	Swarm	3G	Sariyer
11	Test 11	Swarm	3G	Istanbul
12	Test 12	Instagram	3G	ITU Insaat
13	Test 13	Instagram	3G	Istanbul Turkey
14	Test 14	Instagram	3G	ITU
15	Test 15	Instagram	3G	Emirgan Sahil
16	Test 16	Instagram	3G	Sariyer
17	Test 17	Twitter	Wi-fi	with location label (Istanbul, Turkey)

The actions listed in Table 2.2 returned only 10 tweets inserted into database table (Figure 2.1). The first six tweets have not been inserted into this spatial database since they are without spatial component, from 2 to 6 have spatial tags but those are just place names and bring no location information (as a pair of latitude and longitude). Test 7 is tweeted by checking detailed location preference on Twitter and close to the exact user location. On the other hand, while posting on Swarm and Instagram, platforms search places nearby corresponding to GNSS and suggest several places to be attached with the post. These platforms also allow attaching distant locations to the users' exact position by the use of POI library. Therefore, users' have the ability to manipulate their own location. In addition, assumptions in the context of GNSS measurement might disarray data as seen in the rest of 7, 8, 12 test tweets in Figure 2.2 (a) and (b).

test_id integer	twitteruser character varyin	tweet character varying (254)	lat numeric	lon numeric
7	Gulnerman	Test 7	41.10506781	29.01959214
8	Gulnerman	Test 8 (@ n9aat Fak3ltesi - @itu1773 in Sstanbul, T/rkiye) https://t.co/xYbbs400LG	41.10490236	29.01907682
9	Gulnerman	Test 9 (@ stanbul Teknik niversitesi - @itu1773 in rstanbul, Tsrkiye) https://t.co/XGR4JzhzQZ	41.10604004	29.02247441
10	Gulnerman	Test 10 (@ Sar0yer - @sariyerbelediye in Istanbul) https://t.co/ijoAY9JvGJ	41.16469859	29.05694962
11	Gulnerman	Test 11 (@ 1stanbul in Istanbul, Tirkiye) https://t.co/qkNtsnxh1J	41.00479978	28.98244781
12	Gulnerman	Test 12 @ T0 0n:aat Faksltesi https://t.co/Omc8o7voob	41.10479764	29.01899292
13	Gulnerman	Test 13 @ Istanbul, Turkey https://t.co/3CxMVXFnnW	41.0186	28.9647
14	Gulnerman	Test 14 @ stanbul Teknik niversitesi (3Tx) https://t.co/NJCtKXbcbU	41.10328138	29.02333759
15	Gulnerman	Test 15 @ Emirgan Sahil. https://t.co/PMEDPHsm1N	41.10483249	29.05657758
16	Gulnerman	Test 16 @ Sar:yer, .stanbul https://t.co/SaMMoMosCw	41.1667	29.05

Figure 2.1 : Test Tweets Database Table.

The user location is marked with the orange triangle in Figure 2.2 (a)(b). Detail point of the disarray idea is assumption context. All the test data are sufficient to be used for a relatively small-scale project such as tourism or migration inter cities or countries. On the other hand, for some large-scale projects like building and street levels, the data is required to be questioned for this detail of mapping. Tweets can be either assessed, cleaned, relocated or corrected by using clues of content, metadata. Moreover, the platform where a tweet sent from is identifiable with the "https://t.co/..." link in the tweet (Figure 2.1).

The test study shows that the actions with detailed location preference on Twitter and geotagged posts on Instagram and Swarm have captured as geo-referenced tweets. While detailed location preference directly attaches latitude and longitude with the accuracy of GNSS, the other SM sourced locations might be manipulated by the user. Although georeferenced SMD embody the small percentage of all SMD and

incorporates possible spatial disarray, this study carries out that data in order to demonstrate its stronger spatial correlation to the events (Issa et al, 2017).

(a) (b)

Figure 2.2 : Test Tweets; (a) Distribution of Tweets, (b) Distribution of Tweets Near User Exact Location.

2.3.2 Disaster event case

On the 15th of July, 2016 Turkey was exposed to a case of emergency, i.e. an attempt of a military coup. Although the event spread widely in many cities, the vast majority of people first learned about this emergency from SM. In the first minutes of the attempt, the media, including news channels broadcasted nothing. However, SM users raised questions on the sounds of jet planes in the capital city of Turkey and the blockage of strait bridges in Istanbul. While the traditional media stayed silent about the coup and broadcasted ordinary programs, people on SM had already started to discuss about a coup and they drew the attention of the whole world. In this study, the SM reactions to a disaster were assessed by textually and spatially to show how SMD serve as a new age journalist. The following sub sections mention about this serve in the form of this disaster case in the order of; data capture tools, data overview, data process including text and spatial data analysis sub parts.

2.3.3 Data capture tools

There are diverse ways to capture and analyse SMD, a comprehensive collection of toolkits are listed by Ryerson University Social Media Lab (Ryerson University.).

Since the twitter data is in use of this study, prominent toolkits related with twitter data are briefly introduced. The “twitteR” (Gentry, 2012) is an R package to download twitter data by “username”, “keywords”, “time interval”, etc. This option allows a direct R coding environment to data analyst that can be a good start with the statistical features of R. Tweetcatcher Desktop (TCD) (Brooker et al, 2016) is another free desktop program that harvests and visualizes twitter data for social sciences. Knime is a platform to provide ability of tweet harvesting with the Knime Twitter Nodes package. Knime also has SM Sentiment Analysis tools and text processes including NLP, text mining and information retrieval (KNIME, 2014). Carto is an online mapping platform, which provides certain amount of georeferenced SM data as free and allows drag and drop spatial analysis (Carto, 2014). Moreover, there is NodeXL interface to download the data which has basic and pro versions to make text analysis and see the graphs of network between twitter users (“NodeXL for Programmers,” 2011).

In this study, Geo Tweets Downloader (GTD) is used for data collection (Gengec, 2016). GTD has been designed for our previous study (Gulnerman et al, 2016) that serves the ability of defining the spatial location, time and the text of the posts at the same time. With the help of the GTD, an analyst can collect public tweets from Twitter stream in a determined time interval based on a user-defined boundary regarding the research. GTD is a desktop application developed in Java using Twitter4j (Yamamato, 2015) library. Twitter have RESTful API and StreamingAPI for developers to manage, use and query Twitter functions (Twitter, 2020a). With the use of Twitter4j, developers can implement RESTful and StreamingAPI functions of Twitter with their own authentication parameters to their Java applications.

The application listens to stream that consist of instant tweets by twitter users who allow their tweets to be public in their privacy settings. GTD filters the tweets from the stream using Twitter RESTful API and StreamingAPI, which have geographic coordinates in a user-defined MBR (Minimum Bounding Rectangle) boundary. The application sends the filtered tweets to the (PostgreSQL, 2017) database supported with the (PostGIS, 2017) extension. Following this, sent data converted to a point base vector in shp (shape file) format including the attributes of “twitter_username”, “tweet_text”, “tweet_time” and geographic coordinates once the application detects a tweet within the user defined boundary area.

2.3.4 Data overview

Case data is bounded by Istanbul that is the most populated city in Turkey. One of the first acts of the Putschists in Istanbul was the blocking of the first and second bridges over the Istanbul Strait which connect the Asia and Europe continents. This resulted in the news break since more than 360 thousand vehicles are crossing these bridges everyday. From the first tweet at 20:50 on the July 15 until the midnight of July 16, 2016, there were totally 7,883 geo-referenced tweets within the determined boundary. There is no text filtering for this project data capture other than the spatial filter. Bruns and Liang (2012) denotes that the first challenge of SMD capturing during crisis is finding related posts with the crisis and gathering. One approach is as used in several studies focusing on topical #hashtag contained posts which does not mean you gathered all the crises related data (Bruns & Liang, 2012). In this study, there is no direct focus on any #hashtags for assessing and all collected data is evaluated. Moreover, with the knowledge of geo-referenced tweets embody a small part of all tweets on the event time, all geo-referenced tweets, that are positioned either directly by GNSS or indirectly by POI libraries of other platforms, are processed. In this way, the aim of this study is constituted as to evaluate the information of the tweets, which have geographic coordinates considering both the spatial analysis and the text analysis.

Tweet counts from the start to the midnight of 16th July 2016 is displayed in Figure 2.3. The number of tweets steadily decreased from 3am to 8am. After 8am it climbed up to 400 tweets per hour to 13:50 and afterward it slowly fluctuated until 19:50. After 19:50 it continued to rise to 600 tweets per hour.

Tweets trend for coup attempt day has been investigated with the tweets of one week before as same day and same time interval. During the coup attempt day, there are 7883 tweets tweeted by 5336 different twitter users in total while 12033 tweeted by 8180 different twitter users in one week before. There are 20 twitter users who tweeted more than 10 times during coup attempt and the most active user has 39 geo-referenced tweets. While one week earlier, the number of the users who tweeted more than 10 times is 12 and there is one user who has 142 tweets on that time which might cause bias and removed from the dataset. It is clearly stated that the datasets of this study do not include any bots' tweets. Bots, as automated programs, could be legitimate or

malicious which generate large amounts of tweets related with news and feeds or spread spams (Zi et al, 2010).

The given line graph shows the change of tweets count for coup attempt day and the same day one week before over 24 hours. While the total amount of the tweets is 50% more than tweets on the day of the coup, it is the opposite between 01:50 am and 07:50 am. There is no significant anomaly for the rest of the time. This time interval covers the series of events and attacks occurred due to coup attempt.

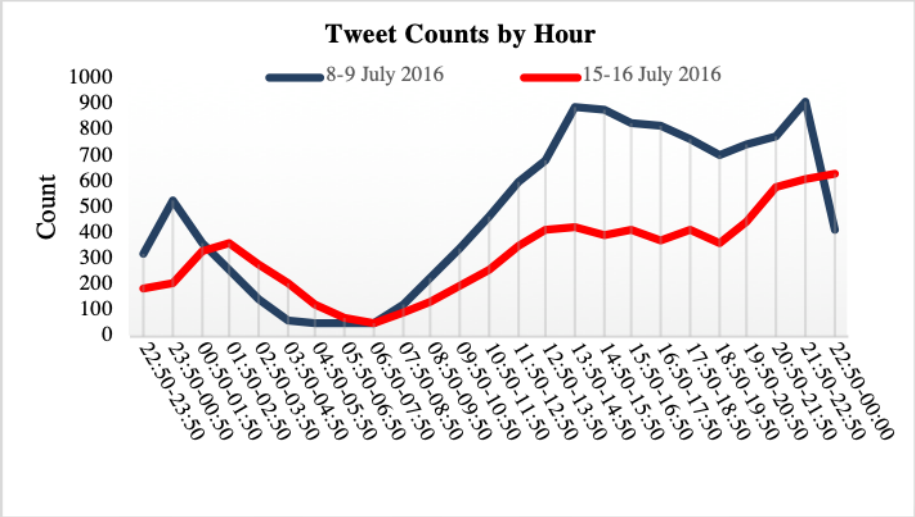


Figure 2.3 : Number of tweets by the hour for the event day (15-16 July 2016) and one week before (8-9 July 2016).

2.3.5 Data process

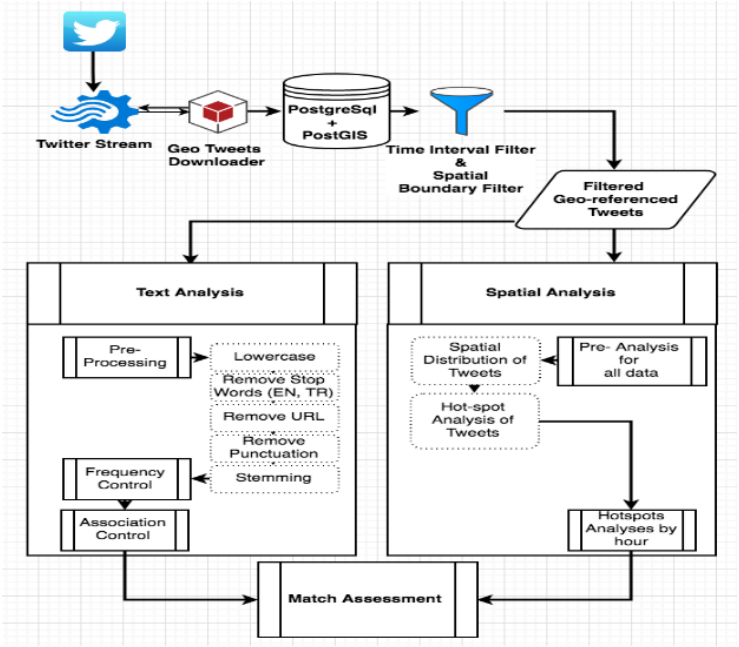


Figure 2.4 : Data Process Flowchart.

Data processing approach is conceptualised in Figure 2.4. GTD captures only geo-referenced tweets and insert them into PostgreSQL database with PostGIS extension. Determined time interval and the spatial boundary are applied to geo-referenced data to filter. Filtered data is assessed with text and spatial analysis, which are detailed in the next subsection.

2.3.5.1 Text analysis

The textual evaluation methodology of this study comprised text analysis to understand the SM dominance topic. The text analysis has three sub-steps; (1) pre-processing steps (Figure 2.4), (2) calculating the frequency of the words, and (3) finding the correlation between words.

R programming language with RStudio is used for the text analysis (R, 2017; RStudio, 2016). RStudio has both open source and commercial integrated development environment (IDE) for R that is a language and environment for statistical computing.

DBI (R Special Interest Group on Databases, 2016), R PostgreSQL (Conway et al, 2016), and R PostGIS (Bucklin and Basille, 2018) packages were used for database connection. R Studio has text mining (TM) package (Feinerer, 2017) that enables pre-processing. The pre-processing includes the removal of white spaces, URL, and punctuations. Besides, short texts also include insignificant words that are called stop words. Since stop words has no discrimination between topics, it is meaningful to remove those words in the pre-processing part. Unfortunately, TM package serves the list of stop words for the English language but not for Turkish language, a curated Turkish stop words list (Aksoy and Ozturk, 2018) is scanned and removed from data.

Figure 2.5 presents the ordinary text cleaning steps; lowercase letters, removing stop words for English and Turkish, removing URL, removing punctuations from the corpus and stemming (Cakir and Guldamlasioglu, 2016; Hearst, 1999). The stemming step, is to find the roots of words, that is useful for removing suffixes and counting similar words in one hand. SnowballC is a stemming package that has a Turkish extension as well (Bouchet-Valet, 2015; Cilden, 2007). However, stemming allows us to match the words without noises of different forms of the same words, like organize, organizes and organizing (StanfordNLPGroup, 2008), Turkish is not an easy language for stemming (Cakir and Guldamlasioglu, 2016; Seker, 2014). Some of the words, like Turkiye (Turkey), is stemmed as “Turki”, meydan (square) is stemmed as “meyda”

and demokrasi (democracy) is stemmed as "demokras" in Turkish as correct as in the first form. To minimise such cases, the last pre-processing stemming step is deliberately ignored.

Figure 2.5 : Pre-processing Steps of Text Mining Flow Diagram.

Following the pre-processing part, to identify the most frequently used words, document-term-matrix (Feinerer, 2017) function within TM package is used. This function categorizes words in the document into columns and the rows are filled logically considering whether or not each term in the document is used. At the end, the sum of the rows for each term gives the number of documents including the term. Therefore, the most frequent words are used to understand what the unconscious volunteers talked about the most within 24 hours. The lack point of this technique, that would affect the result is there is no synonym or similar meaning control because the text frequency just counts the exactly the same words. The top fifty the most frequently mentioned words are shown in Figure 2.6., Turkey and Istanbul, have been omitted for obvious reasons. There is a big difference between these two words and the third and the rest of the fifty words. Therefore, these words were counted as outliers to show on the bar plot and to make associations with other words.

The 50 top words list includes several district names of Istanbul and some strategic locations such as; airport names ("*atatürk*", "*sabiha*", "*gökçen*" and "*tavairports*"), bridge names ("*fatih*", "*sultan*", "*mehmet*" - "*boğaziçi*"). The rest of the words are reflecting the re related to disaster as; "*vatan*" (motherland), "*meydanı*" (square), "*darbe*" (coup), "*asker*" (soldier), "*bayrak*" (flag). Associations of these words may give us a better and systematic perspective as to what they really are instead of being estimated by an author's individual perspective. The most frequently used word was "*vatan*" (motherland), after the three districts names of Istanbul; *kadıköy, bakırköy* and *pendik*, which is associated more than 20% with "*bölünmez*" (indivisible), "*feda*" (sacrifice), "*bayrak*" (flag), words. The word "*meydanı*" (square) was 48% associated with "*taksim*" which indicates the heart of the city "Taksim Square". The word "*havalimanı*" (airport) and was 74% correlated with "*atatürk*", 70% with "*tavairports*", 55% with "*uluslarası*" (international), 46% with "*gökçen*", 46% with

"*sabiha*", 44% with "*bakirköy*", which are all defined as the two airports of Istanbul; "*Atatürk*" and "*Sabiha Gökçen*". In addition, "*sabiha*" was 99% associated with "*gökçen*" which shows they are a completion of each other. Another word "*köprüsü*" (bridge) was 65% correlated with "*boğaziçi*", 55% with "*mehmet*", 46% with "*sultan*", 41% with "*fatih*" and specified the "*Boğaziçi*" and "*Fatih Sultan Mehmet*" bridges. Moreover, the correlations between "*mehmet*", "*sultan*" and "*fatih*" are higher than between "*köprüsü*" (bridge) which explains the completion of words. The word "*darbe*" (coup) and was 27% associated with "*girişimi*" (attempt) that clearly informed about a "coup attack". The last word for this text study is "*bayrak*" (flag) which was 65% associated with "*dinmez*" (inconsolable) and 62% with "*inmez*" (not haul down), which shows the reaction of Turkish citizens living in Istanbul.

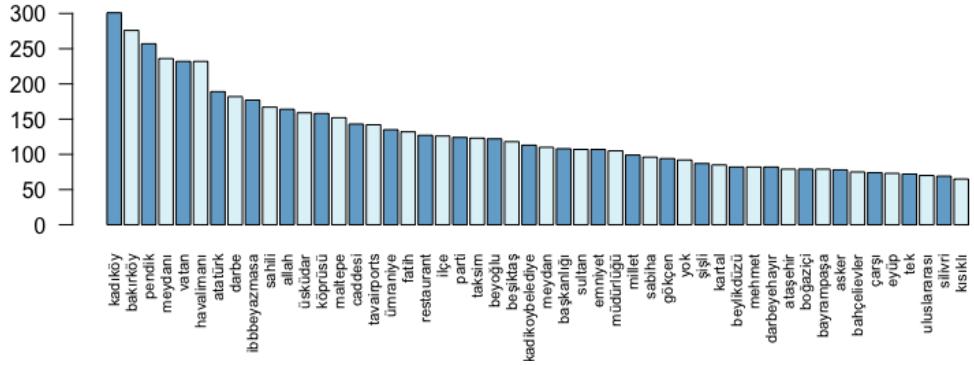


Figure 2.6 : The most frequent terms in the Tweet dataset within the defined time interval.

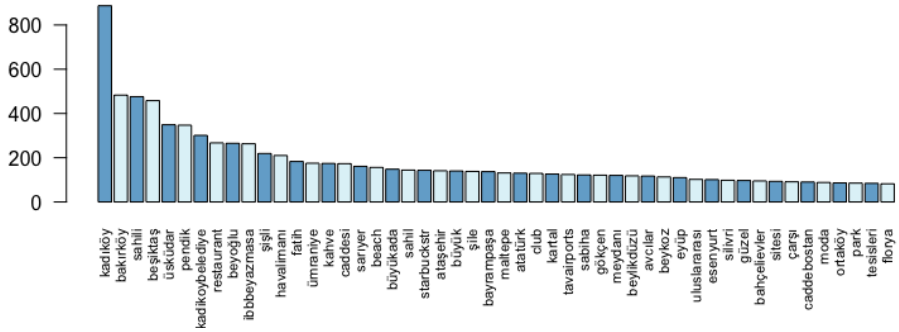


Figure 2.7 : The most frequent terms in the Tweet dataset one week before the defined time interval.

On the other hand, one week before the defined time interval, the top frequent words have been displayed in the Figure 2.7. It is clearly stated, most of the districts like

"*kadıköy*", "*bakırköy*", "*pendik*", "*üsküdar*", "*maltepe*", "*ümraniye*", "*fatih*", "*beyoğlu*", "*şişli*", "*kartal*", "*atasehir*", "*bayrampaşa*", "*bahçelievler*", "*eyüp*" and "*silivri*" and strategic locations names also were in the top 50 frequent words. However, the rest of the frequents words are "cafe", "restaurant", "*kahve*" (coffee), "home", "starbuckstr", "club", "park" that are mostly looks like a Friday night and weekend activities on Saturday. Moreover, there is no tweets containing "*bayrak*" (flag), "*asker*" (soldier), "*darbe*" (coup), three tweets contain "*vatan*" which are featured words within the coup day tweets dataset.

2.3.5.2 Spatial data analysis

Spatial analyses are to gain a general perspective of data sprawl over the defined area. They result in a spatial distribution of tweets by hour for 24 hours period, the analysis of spatial distribution for one week before over the same spatial boundary. Same time interval by three hours and the comparison assessment between the coup day analyses and the analyses one week before. Spatial distribution of tweets by hour is given in Figure 2.8, demonstrating the hourly change of the number of tweets and sprawl over Istanbul. As it is shown, the non-urbanized parts of Istanbul, have not much data at almost any contributed in terms of time intervals. Also, the changing spatial sprawl of data, could indicate two things: (1) there would be the emergency situations at the same location of the tweets or (2) the reactions of volunteers locating different places. In addition, the repetitiveness and intensity of the data sprawl. They seem interesting may be associated with the rise of new breaking events or the reduction of event effects. In fact, on the July 15th the crises were composed of the numerous attacks at different location and time. Some of them were instantaneous or short-lived events, like low altitude flights of jets making sonic bombs and the conflict between police and soldiers, and non-instantaneous events, like the closure of strategic points of transportation and long-term television broadcast interruption.

All these extractions explained above are conceptual, however, to identify more precise information extraction from spatial data, hot spot analysis was tested over hourly spatial datasets. Optimized Hot Spot Analysis (ESRI) was used within ArcGIS software, which used the Getis- Ord Gi* statistical technique. This technique unlike the Moran I index and Geary C ratio techniques, provide with a presentation of the places of clusters (Cubukcu, 2015). Within the optimized hot spot analysis, the

incident data aggregation method is chosen for "count the incidents within fishnet polygons" for defining the number of incidents within each polygon. In addition, the bounding polygons are defined as the Istanbul city boundary to compare the results of each cell, even if the cells had no incidents for each time interval. Furthermore, there was no analysis field selection to process attribute data, since all the data weights are assumed to be equal (ESRI).

The aim of the Getis-Ord Gi* statistical technique is to differentiate the clustering pixel size based on the data sprawl and density. Although the tool Optimized Hot Spot Analysis allows to analyze pixel size manually, making all pixel size equal for all-time interval may provide a lower clustering capability and potentially lower clustering accuracy. Therefore, it is better to choose Getis- Ord Gi* technique with its defining capacity of pixel size for each time interval dataset.

Optimized hot spot analysis results provide four attribute columns, which are the value of the confidence interval of each cell (z), the rest part of the confidence interval (p), the value for a cold spot, insignificance and hot spots (Gi-bin) and the number of points for each pixel (join count). Hot spots analysis of each time interval was classified into three classes, corresponding to their own natural break intervals. The darkest red pixels showed in Figure 2.8 correspond the areas with the majority of the tweets as clusters that may have been assessed with the emergency.

Hourly hotspots of the Istanbul boundary provided a more detailed perception about which areas were denser, seen in Figure 2.8. The size of pixels varied between the hourly analyses and provided two pieces of information. If there was a more dispersed data cluster, there are the smaller sized pixels that offered a more precise location of events. Otherwise, a bigger pixel size may indicate superclusters or less information enriched sites. So, the total number of tweets count should be considered while interpreting the data. In the context of the views above, the zero-time interval, darkest red areas, had dominancy compared to its consecutive, as interpreted as the first perception of systematic attacks at many strategic points of the city. Within the zero-time interval, Bosphorus Bridge and Fatih Sultan Mehmet (FSM) Bridge had been closed down by soldiers, Ataturk Airport and Sabiha Gokcen Airport were occupied by soldiers and air traffic was stopped and more jets started to fly at the low altitude and tanks and armoured personnel carriers had also been observed. All these unusual events attracted SM attention in the zero-time interval, as seen in Figure 2.8.

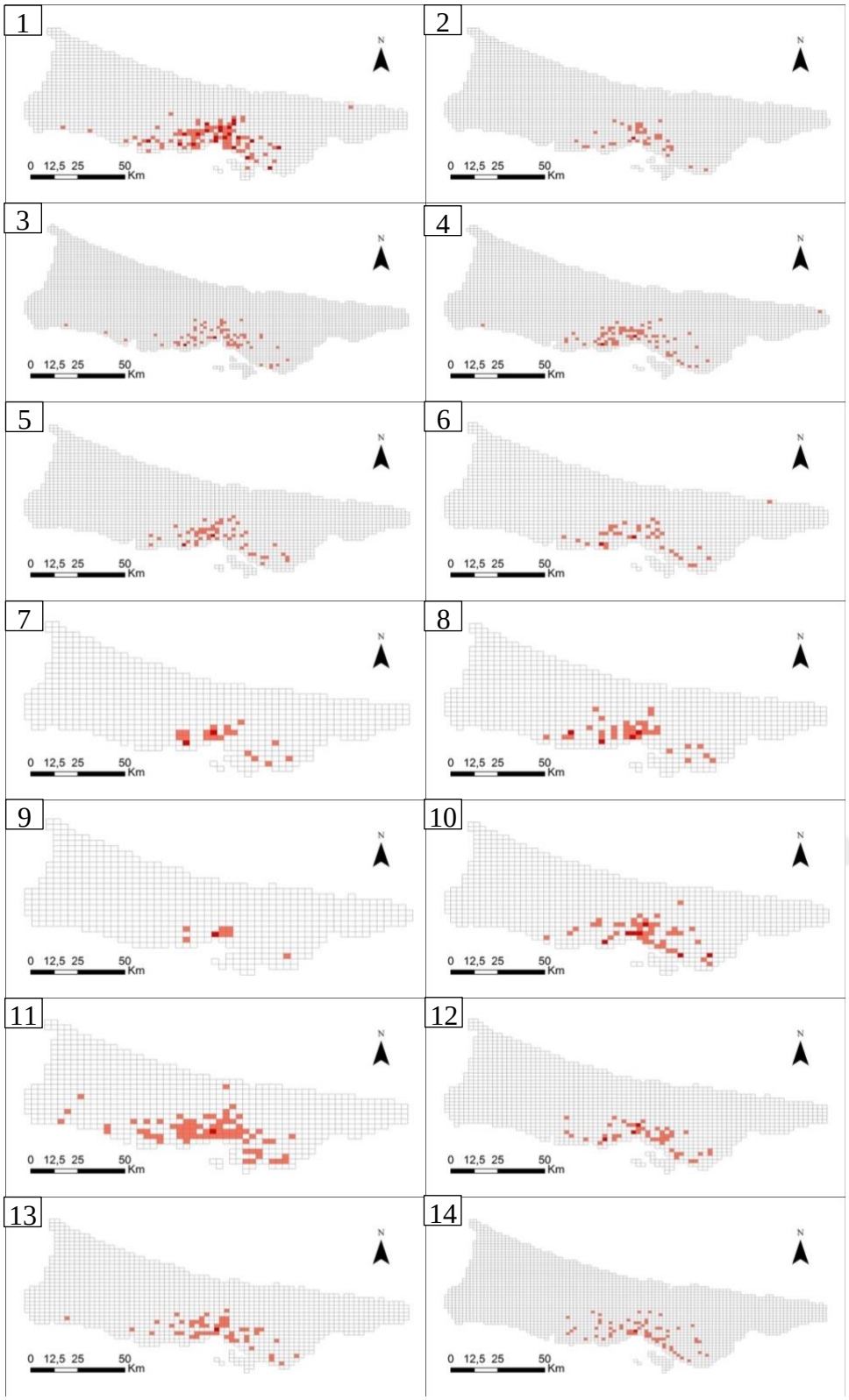


Figure 2.8 : Hotspots of tweets - spatial distribution for 24 hours.

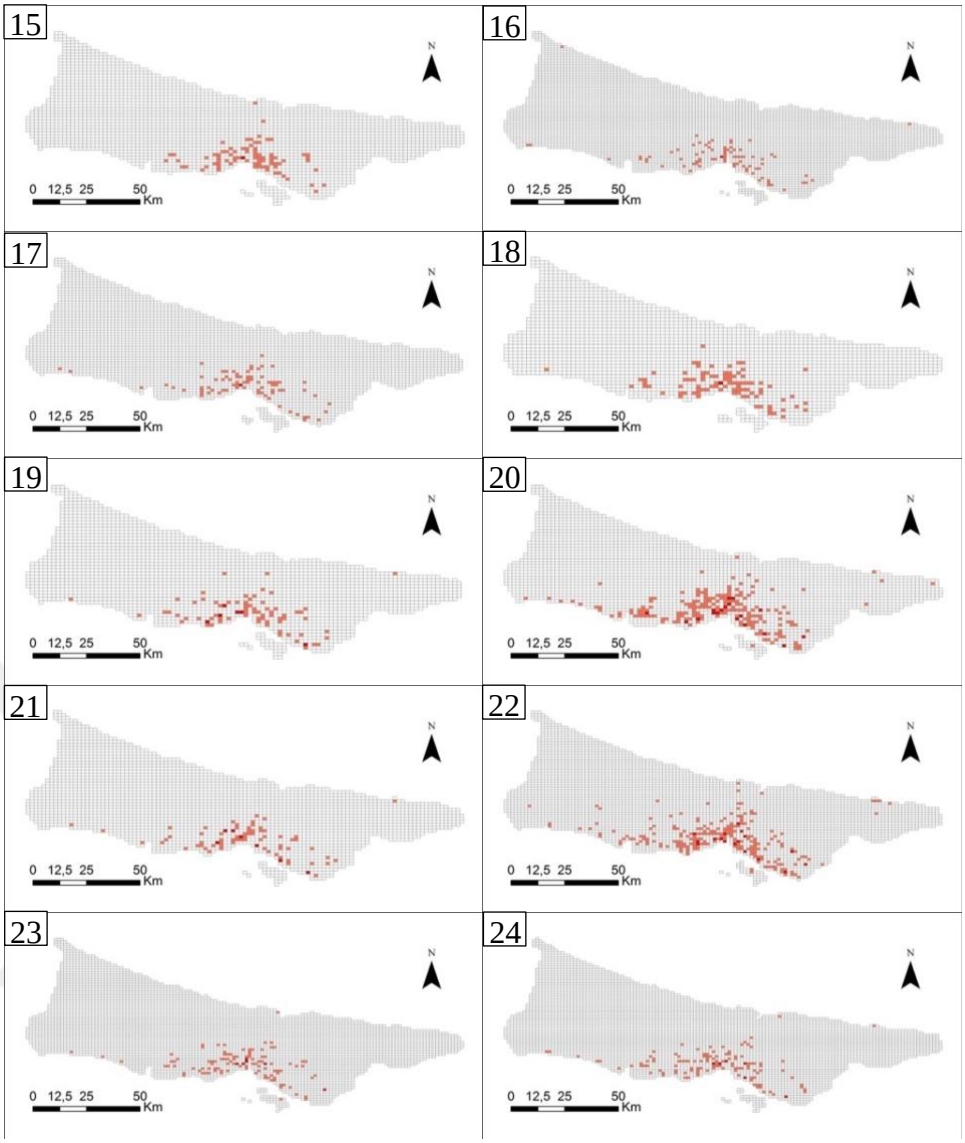


Figure 2.8 (continued) : Hotspots of tweets - spatial distribution for 24 hours.

All of the later time intervals had different hotspot areas, distributions and the number of hotspots pixels can vary. In some of the time interval analysis, there were many hotspots spread within the spatial boundary, like 1st, 20, 22, while some of them such as in 7th and 9th time intervals are more focused on one small or precise location. In some of the time interval analyses, there was also the dominance of interim spots with its hotspots, especially in the 1st, 3rd, 4th, and from 14th to 24th. However, in the 11th time interval analysis, there was just one hotspot pixel with many interim pixels.

The size of pixels can vary based on the count and the spatial distribution of data (Figure 2.8). The reason behind that could be the increase in the number of events in different locations. While analysing the hourly changes of an area, the trend of the actual area should be considered to avoid having a biased judgement. The area may

have dominant events for the similar day or night time in the similar spatial area. In this regard, the data of an earlier week for the same spatial boundary have been studied.

Human activities on SM platform have their own spatial and temporal patterns within normal days. For example, the number of tweets has its own patterns at different hours of days of the week. The analyses on MapCalc software (Red Hen Systems), showed how the temporal pattern during the disaster differ from the daily pattern, and displayed the spatial distribution of people during the disaster differ from normal days.

According to the intelligence reports on the newspapers beginning of the event was taken as the first-time interval, the correlation coefficient is determined as 0.695 almost highly correlated with the ordinary days' same time interval, during the expansion of the events, when most of the people are not aware of what was going on, which is second time interval, the correlation is still high according to the correlation coefficient which is 0.716. However, when the case event reach to the highest action time of the attacks where the time intervals are 3, 4, and 5, the analyses show that the correlation coefficients are low enough to prove the tweet pattern of the people is changed drastically. The correlation coefficients are determined as, -0.016, -0.003, and 0.004 for the third, fourth, and fifth time intervals, showing the tweet patterns are not the same with the ordinary times.

While results of correlation between a week before the attack and attack day dataset for the first-time interval is 0.695, the correlation is loosening as it goes further, e.g. for the fifth-time interval it is 0.004. The fifth-time interval coincides between 02:50 am and 03:50 am. This could be due to sleeping time as normally majority of people are sleeping in these times instead of tweeting or reacting to attacks. It is clearly stated that the results of correlation between datasets one week before and the attack day is decreasing considering the time-interval because attacks have been lastlong actively until morning. This is also interesting to justify why we need to capture and analyse the normal days data as it may give a better understanding of important intervals (e.g. sleeping period).

Moreover, the other outcome of analysis result; coincidence table explains the details of difference. The table displays comparison between value of cell through number of cells, number of cross-cell, % of cells' area, average %, standard difference and weighted score. According to the table, standard difference clarifies the change

between two maps and in the hot spot values proves the change with standard difference varying between 1.97 and 2.0 except for the highest area cover group that reflects general trend in the spatial boundary.

2.4 Validation & Assessment

The last part of the study is about the comparison between what SMD reported and what the actual events happening. This is particularly important for credibility and reliability analysis of using SM for event detection. The events occurred for during 24-hour time interval were searched in the traditional media as published. Our traditional media sources were websites of newspapers, online sources, TV channels and independent news web pages.

Table 2.3 : Event locations and counts in the July 15 coup attempt (Gulnerman and Karaman, 2017).

Event Location	Event Count
Bosphorus Bridge	7
Ataturk Airport	6
Istanbul (related the whole city)	5
Harbiye	4
Fatih Sultan Mehmet Bridge	3
Bagcilar	3
Sabiha Gokcen Airport	2
Taksim	2
Istanbul Strait	2
Provincial Directorate of Security	1
Bayrampasa Action Force	1
Fatih	1
Kadikoy	1
Maltepe	1
Uskudar	1
Total	*39*

The SMD validation for this study is done by comparing the results of text analysis and spatial analysis with the news published by trustworthy and credible such as National Newspapers and their official websites. The two instances for our media source in Figure 2.9, are a newspaper named "*Hürriyet*" (Hürriyet, 2016) and a TV channel named "*NTV*" (NTV, 2016) websites which are both well-known, old and rooted media sources in Turkey. The study has only filtered the events that have time and location information like geotags as illustrated in Figure 2.9 (a) or georeferenced and labelled on the map as shown in Figure 2.9 (b). There were 39 events, which

occurred up to 11 hours after the start time of the coup attack (Gulnerman and Karaman, 2017).

Considering those 39 events temporally and spatially defined in the traditional media sources within our past research (Gulnerman and Karaman, 2017), the location of events has been summarized and sorted according to counts in Table 2.3. Considering that, during a 12-hour time interval, the SM reaction was associated with the locations of the hotspots; however, the data that we used in our study was for a 24-hour time interval, which covered a longer period. As well as this, the results of text analysis correlated with the event locations, which were extracted from traditional media (Gulnerman and Karaman, 2017).

(a) (b)

Figure 2.9 : Traditional Media News in online; (a) News in a Newspaper Website and (b) News in a TV Channel Website (Hürriyet, 2016; NTV, 2016).

2.5 Conclusion

Data sources for disaster management, especially in a time of emergency, have various channels. Expanding views of emergency mapping have enhanced the capacity building of resilience by the support of the "United Nations Platform for Space-based Information for Disaster Management and Emergency Response" (UN-SPIDER) program, since 2006. The main aim of the organization is to provide all types of space-based information usage for the whole disaster management cycle (Boccardo, 2013). On the other hand, since 2012, GMES, GIO-EMS services in Europe have tried to support emergency management by earth observation satellites. Although, these initiatives have helped to the management of several environmental disasters, the question is "Could they serve every disaster?". The question is not just limited to the restriction of both vectorization and the attributes of event locations and conditions,

but also is restricted to the question “Could every emergency be sensed by satellites?” (Goodchild, 2007a). There are disaster events like terror attacks in our case that cannot be sensed by remote sensing satellites.

Time management and the potential of response capacity building during the crisis are more important than anything, especially with scarce emergency resources, security forces, etc. The volume of SMD and count of active users have been increasing day-by-day (Statista, 2017b) and this has already started to serve as data source for emergency management (Ao et al, 2014; Power et al, 2015; Yin et al, 2012) from several different aspects and has the potential to serve much more than implemented studies (Houston et al, 2015). While the data volume has been increasing, the process of extracting meaningful, precise and reliable information from the data is getting harder. Extracting information from this unstructured raw data stream requires pre-processing steps to clean and to test the reliability of the data, and processing steps to cluster and classify the data (Wang et al, 2016). All SMD require a validation and accuracy test because of the unknown source of data that is provided by unconscious volunteers. However, Leetaru et al. (2013) has noted that authorities are using Twitter during major disasters to provide real-time information and directives to disaster victims while the emergency responders are initially trying to monitor Twitter as a parallel 911 emergency call system. That monitoring especially required where the telecommunication services are unavailable (Khorram, 2012; Yin et al, 2012). Because of that, our further study will be on spatial reliability and textual reliability to have understanding of which data requires validation or directly used.

The most crucial part of using SM for emergency situations is monitoring by geo-referenced posts (Leetaru et al, 2013) which enabled around 2010 both for Twitter and Facebook (Gross and Hanna, 2010; Moffitt, 2017). However, SM geo-referenced data embodies the small amount of all data (Issa et al, 2017; Power et al, 2015), there are several studies to enlarge geo-referenced data volume by geo parsing techniques (Gelernter and Balaji, 2013; Gelernter and Mushegian, 2011; Lee et al, 2013) and to demonstrate the reliability of the geo-referenced data comparing the non-geotagged data (Issa et al, 2017). In this study, datasets are composed of just geo-referenced data and none of geo-parsing methodologies have been applied considering geo-referenced usefulness. However, to extend the data volume in further studies, geo-parsing techniques will be considered.

VGI is seen an inexpensive, wide and sometimes quick way of data gathering; however the VGI quality is a debate regarding several conceptual quality assessments such as accuracy, granularity and consistency (Mocnik et al, 2018). While researchers are studying as a peer on it, the others are claiming the data has full of unreliability (Fonte et al, 2015; Haklay, 2010). The reliability discussions and data cleaning perspectives are varying for SM-VGI. Some of them are trying to eliminate the reliability according to discriminate the users by using profiling (Abbasi and Liu, 2013), some of them are categorizing the accounts as human, bot and cyborg (Zi et al, 2010). Although, within geo-referenced SMD studies includes several outlier removals in the spatiotemporal context (Bassolas et al, 2016; D'Andrea et al, 2015; Hasby and Khodra, 2013; Lenormand et al, 2014), there is no such a study directly focusing on the ways of cleaning and testing the reliability of the spatial data gathered from SM. Within our study, there is some cleaning actions and filtering have been implemented within pre-processing steps, however deeper cleaning and filtering techniques would be used for our future studies.

In this study, the VGI concept overviewed with its place in the literature and the sub three title of VGI are explained. Moreover, as a sub title of VGI, the contribution of SM-VGI to disaster management mentioned and exemplified with various studies. In addition to that, a case study carried out as an example of series of emergency event. And the SM reaction against those emergency situations analysed textually and spatially. There are some inferences from the case. The one is that simply processes without deep analysis techniques of topic modelling and spatial statistics, SMD would provide valuable information to monitor events. The second, SMD needs to be assessed corresponding to number, duration, repetitiveness and intensity of events for better understanding of spatial distribution of reactions in time. The third, SMD requires proper validation and reliability techniques to contribute emergency management. Further studies would include the events sequential logic and the reliability of the data considering cause and result relation.

3. SPATIAL RELIABILITY ASSESSMENT OF SOCIAL MEDIA MINING TECHNIQUES WITH REGARD TO DISASTER DOMAIN-BASED FILTERING[2]

3.1 Abstract

The data generated by social media such as Twitter are classified as big data and the usability of those data can provide a wide range of resources to various study areas including disaster management, tourism, political science, and health. However, apart from the acquisition of the data, the reliability and accuracy when it comes to using it are concerning scientists in terms of whether or not the use of social media data (SMD) can lead to incorrect and unreliable inferences. There have been many studies on the analyses of SMD in order to investigate their reliability, accuracy or credibility, but not dealing with filtering techniques associated with the data before creating the results, or after their acquisition. This study provides a methodology for detecting the accuracy and reliability of the filtering techniques for SMD, and then a spatial similarity index that analyzes spatial intersections, proximity, and size, and compares them. Finally, we offer a comparison which shows the best combination of filtering techniques and similarity indices to create event maps of SMD by using the Getis-Ord Gi* technique. The steps of this study can be summarized as follows: an investigation of domain-based text filtering techniques for dealing with sentiment lexicons, and machine learning-based sentiment analyses on reliability and developing intermediate codes specific to domain-based studies. Then, by using various similarity indices, the determination of the spatial reliability and accuracy of maps of the filtered social media data. The study offers the best combination of filtering, mapping and spatial accuracy investigation methods for social media data, especially in the case of emergencies, where urgent spatial information is required. As a result, a new similarity index based on the spatial intersection, spatial size and proximity relationships is introduced to

[2]This chapter is based on the article: Gulnerman, A.G., Karaman, H., 2020: Spatial Reliability Assessment of Social Media Mining Techniques with regard to Disaster Domain-based Filtering, ISPRS International Journal of Geo-Information, 9, 245. doi: https://doi.org/10.3390/ijgi9040245

determine the spatial accuracy of the fine-filtered SMD. The motivation for this research is to develop an ability to create an incidence map shortly after a disaster event such as a bombing. However, the proposed methodology can also be used for various domains such as concerts, elections, natural disasters, marketing, etc.

Keywords: VGI; spatial assessment; spatial similarity index; sentiment analysis

3.2 Introduction

This study focuses on finding an assessing methodology to quantify impacts of filtering techniques on spatial reliability of the social media data (SMD) to acquire the most suitable, reliable and accurate SMDs based on the subject for using it in various approaches. The contextual errors at the filtering stage of dealing with SMD result in focusing on irrelevant events, irrelevant reactions and irrelevant locations with regard to the event. These errors result in the creation of unreliable or inaccurate maps relating to the event - what we will refer to as the domain in this study. Many studies have examined the reliability of SMD mostly that which is found on Twitter (Chen et al, 2015; Gupta et al, 2013; Wang and Zhuang, 2018). The Twitter platform is selected as the SMD resource for this study due to its geotagging options and wide usage throughout the world (Statista, 2017c).

Social media has diverse topic content sourced by human sensors (Houston et al, 2015). However, it does not provide a direct motivation to contribute to data production apart from the existence of a number of volunteer-based data gathering platforms (Scistarters, 2017; Ushahidi, 2017; Zooniverse, 2017). This makes social media analysis different from the analysis of structured data. Therefore, there is a need to filter such data to retrieve the relevant data for a selected domain.

For coarse-grained hot topic analysis, for example, if there has been a terror event in a city, non-relevant outliers might easily be filtered out from streaming data by using density-based clustering approaches (Ozdikis et al, 2017; Tamura and Ichimura, 2013). Conversely, fine-grained filtering can robustly discretize noisy content for fine-grained analysis (Middleton et al, 2014). For instance, incidence mapping during and following a terror event, and the fine removal of discrepancies by fine-grained filtering, helps to produce more reliable maps to detect the location and the impact of

the event. This in turn can support coordinated and accurate responses and the management of the emergency in question.

The terms accuracy, reliability and credibility are generally used to evaluate the quality of SMD. The relationship between spatial information extraction and the truth in social media analyses is defined as its accuracy by (Achrekar et al, 2011; Chen et al, 2016; Ilina et al, 2012; Ryoo and Moon, 2014; Sadilek et al, 2012). These studies used the concept of accuracy for the estimation of the living, working, or traveling location of the users based on their social media feeds by using the distance between geo-tagged and actual locations. The distance used in these studies is the Euclidean distance, that is the distance between two points either on a flat plane or the 3D space length that connects two points directly using Pythagoras' theorem (Danielsson, 1980). The concept of reliability mostly focuses on the internal accuracy of the data itself, or the method that is used to process the data as mentioned by (Lawrence, 2014). This involves the investigation of the relationship between two different sentiment analyses statistical scores. Credibility is a term for the accountability of the source of the news in the new and traditional media. For social media credibility involves a consideration of the profile of the user, friendship networks and the acts of the users such as tweets, retweets, likes, and comments as suggested by Castillo et al. (2013). Abbasi and Liu, (2013) considered the credibility of social media by remarking on the activities of coordinated users that act together as a pack.

To date there have been several studies on the credibility (Abbasi and Liu, 2013; Castillo et al, 2013), accuracy (Achrekar et al, 2011; Chen et al, 2016; Ilina et al, 2012; Ryoo and Moon, 2014; Sadilek et al, 2012) and reliability (Ceron et al, 2013; Deshwal and Sharma, 2016; Lawrence, 2014) of SMD but, those are somewhat limited because they have only focused on the results of the analyses, not the filtering algorithms. However, most of the time, SMD are filtered, based on the words and hashtags based on the event before they are analyzed (Crooks et al, 2013; Lin and Margolin, 2014; Murzintcev and Cheng, 2017; Signorini et al, 2011). Considering the negative nature of the disaster domain content, sentiment analysis can be seen as an approach that can be used to access related content from a wider perspective than results from hashtag and bag of words filtering. So, focusing only on the SMD analysis results by ignoring the filtering techniques misdirects the assessments. Depending on the degree of

accuracy and reliability, which were referred to as aspects of the quality (Lang and Wilkerson, 2008) of the SMD filtering algorithm, the quality of the analysis changes.

There are two prominent approaches in the literature with regard to sentiment analysis (Ceron et al, 2013; Deshwal and Sharma, 2016; Nielsen, 2011). The first method utilizes a subjectivity lexicon, involving a glossary with sentiment scores or labels for each word or phrase. It involves scanning documents or phrases to determine the constituent total word score in terms of polarity. The second is a more statistical approach that exploits learning-based algorithms. However, it does not work well in the case of fine-grained phrases such as those found in social media content (Dehkharghani et al, 2016). In addition, statistical approaches might not work well on agglutinating and morphologically-rich languages such as Turkish, Korean or Japanese (Kaya et al, 2012). There have been several attempts to perform and test polarity analyses in the Turkish language, some of which are based on translating English subjective lexicons into Turkish (Aytekin, 2013; Vural et al, 2012), while others rely on classifiers across conceptual topics (Kaya et al, 2012) or rely on the linguistic context in Turkish (Erogul, 2009). The use of sentiment analysis can vary according to the domain (e.g. specific topics such as disaster, opinion columns, or music) that it is used for. While the domain-independent lexicons provide fast and scalable approaches for general purposes, domain-based lexicons are valid across specific topics and cultures (Dehkharghani et al, 2016). In terms of the language used, the richness of the subjective lexicon is as important as the techniques and methodology selected. There are plenty of different lexicons for English with rich word content that quantify polarity by scoring (Nielsen, 2011), by assigning several emotions to each term (Cambria et al, 2012; Mohammad and Turney, 2013), by labelling terms as positive/neutral/negative (Baccianella et al, 2010; Liu and Zhang, 2012) and by scoring each term's polarity strength from minus to plus (Nielsen, 2011). On the other hand, most other languages, including Turkish, lack comprehensive subjective lexicons. However, to the best of our knowledge, there are a couple of lexicons for the Turkish language. (Dehkharghani et al, 2016) generated a SentiTurkNet (STN) lexicon which is equivalent to the English translation of SentiWordNet (Baccianella et al, 2010), while (Ozturk and Ayvaz, 2018) produced a Turkish lexicon consisting of over 5,000 terms, determined as being terms found in frequent daily use, which are labelled as positive, negative or neutral. This lexicon is

entitled The Lexicon of Ozturk and Ayvaz (LOA) in this study. Previous studies of Turkish content involve domain-independent sentiment analysis and is mostly implemented over long texts. Hence, the performance of previous studies is unknown in terms of how they deal with short texts from social media with regard to a domain.

The novelty of this study is the development of a methodology for filtering the SMD based on relevancy and spatial accuracy, by comparing the current techniques on filtering. Another outcome of this study is that we have developed a spatial similarity index to validate the spatial accuracy of the methodology along with the filtering techniques. The SMD of this study is acquired from two terror events that occurred in Istanbul. Consequently, the textual data of the tweets is in Turkish. Another distinction of this study is creating a methodology for a language that is not English. Most of the studies for filtering and sentiment analyses of the tweets relate to the English language (Nielsen, 2011; Terpstra et al., 2012; Vo and Zhang, 2015). It has been discovered that using these techniques without proper adjustments based on the event or language, provides a very low rate of success for agglutinative languages like Turkish due to the use of suffixes at the end of words and large amounts of homonyms (Aytekin, 2013; Castillo et al, 2013; Dehkharghani et al, 2016). As Castillo et al (2013) indicate in their study, filtering and the credibility of tweets in Spanish requires manual labeling due to the possibility of non-relevant classifications.

Those type of situations with regard to filtering and language affect the accuracy and reliability of the resulting maps (event, hazard or risk maps). Previous studies have focused on where the event occurred geographically and how big it was, without considering the filtering techniques used (Chen et al, 2015; Gupta et al, 2013; Wang and Zhuang, 2018). However, the quality considered in this study is based on how correctly the SMD is filtered and how correctly this filtered data is reflected in the creation of maps, both spatially and sentimentally.

In this study, the domain selected is that of terror attacks. The events chosen were two such attacks in Istanbul in 2016, and the all the tweets were acquired in relation to these attacks were in Turkish. To investigate the effects of the filtering accuracy on the maps in geographical terms, the results of every filtering technique was mapped using a generic method. The determination of the accuracy of the filtered tweets involves the use of manually-labeled tweets as ground truth, such as Gupta et al. (2013)

and Castillo et al. (2013) used in their research. Then the maps associated with each filtering technique were analyzed in terms of the ground truth map.

At this stage of the study, various similarity indexing methods were used, and a new similarity index was introduced as the Giz Index. The Giz Index is designed to check the similarities of the values, sizes, and proximity of spatial objects, which were filtered and mapped using a technique involving the success percentages with respect to a ground truth map. The difference of the newly developed similarity index is its ability to provide spatial intersection, proximity and size together. For this reason, it can also be used for many studies to detect the spatial accuracy of the generated, estimated, simulated or predicted maps with the truth. Finally, the results of every similarity index method including the newly developed Giz index, were compared with the ground truth map. The comparison showed the best combination of filtering technique and similarity indices to create event maps with regard to SMD. This study provides a methodology which can be used to accurately and reliably filter SMD, and offers a method for investigating the effects of filtering techniques on map accuracy.

3.3 Materials and Methods

The methodology of this study begins with an exploratory analysis of rough filtered Twitter data to create an approach for fine filtering domain-based data. In this way, noisy outliers which are not related to the domain, could be discretized. To build up the approach for filtering out discrepancies using sentiment analysis and machine learning techniques, the basic workflow of data science (Wickham and Grolemund, 2016), which is import → tidy → understand (visualize → model → transform) → communicate, is followed as shown in Figure 3.1. Similarly, data processing is interpreted by considering the taxonomy of data science (obtain → scrub → explore → model → interpret) as suggested by (Mason and Wiggins, 2010).

Figure 3.1 presents the parts of the methodology in the form of the data import and retain part as shown in white, the data tidying part as grey, the data exploration part as blue, the data processing part as green, and the data interpretation part as yellow, pink and turquoise. Each part of the methodology is schematically extended with the use of figures (Figures 3.2 to 3.7) and explained in the following subsections. In respect to this, subsection 3.3.1. provides a technique for geo-referencing data acquired from the Twitter stream by using Twitter API. This part is compiled with the use of Java as

described in Gulnerman et al. (2016) as Geo Tweets Downloader (GTD). In subsection 3.3.2, the data tidying details are given in the order of pre-filtering, cleaning, and labelling to generate ground truth data. In subsection 3.3.3, data exploration techniques such as word clouds, comparison clouds, and dendrograms that are used for this study are explained. This exploration contributes to the development of the text pre-processing function that is part of the data tidying part.

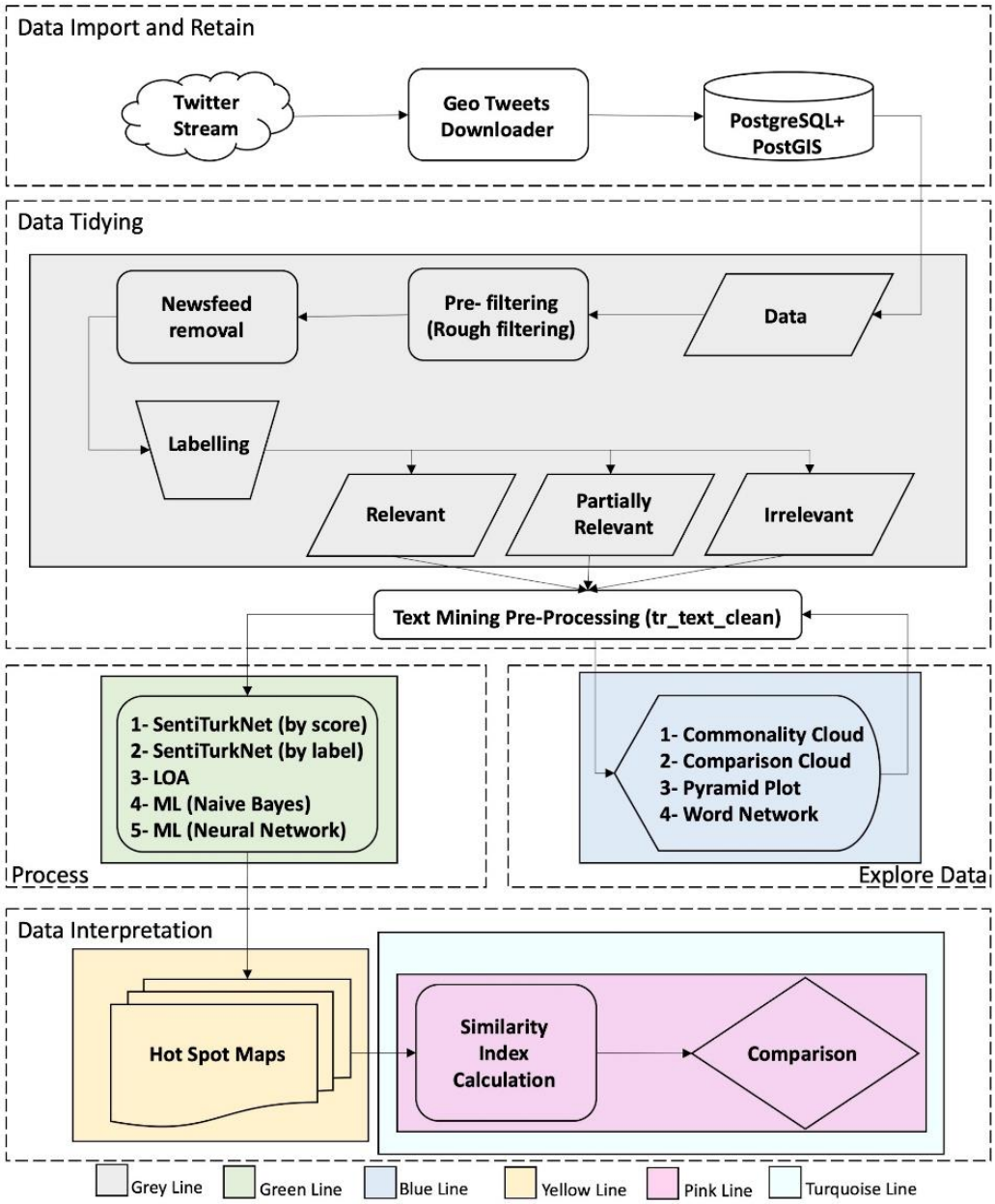


Figure 3.1 : Data workflow.

In subsection 3.3.4, the adaptation methodology of the most common text classifying techniques to a terror domain-based filtering is introduced. In this part, two different sentiment lexicons for the Turkish language and three different machine-learning

techniques are presented for automatically filtering relevant content. In subsection 3.3.5, the spatial interpretation methodology is introduced for quantifying how textual filtering exactly influences the spatial accuracy of the maps produced. This interpretation involves the following steps: 1- production of a hot spot map for manually labelled relevant data (the ground truth map), 2- production of hot spot maps for each filtering technique in terms of relevant outcomes (a predicted map), 3- quantifying the similarity between ground truth and predicted maps (spatial accuracy measure). In this similarity quantification process, current similarity indices (3.3.6.1) and a new similarity coefficient entitled the Giz Index (3.3.6.2) are utilized for quantitatively interpreting the spatial accuracy of the maps produced. Features such as spatial proximity and spatial cluster size are taken into account by the Giz Index but not by the current indices. This is explained and compared with the test data in subsection 3.3.6.2.

3.3.1 Data importing and retaining

Geo Tweets Downloader (GTD) (Gengec, 2016; Gulnerman et al, 2016), a Java-based desktop program that allows filtering by the use of a spatial bounding box and omits non-geotagged tweets, is adopted to collect and spatially filter Turkish data. GTD utilizes Twitter APIs (Twitter, 2020a) that provide public status tweets in real-time. It also helps configuration with regard to the use of PostgreSQL in order to store retained geotagged data. By means of GTD, public status data is collected continuously up to the limits of the API usage. Since the tweets collected are of public status and geotagged, the number of tweets that were posted during the period under consideration was lower than normal. However, the aim and technique used does not need to have all the tweets posted to test the approach which involves highly related filtering.

3.3.2 Data tidying

Social media analysis with regard to an event starts with accessing the related data chunk. This is mostly obtained by querying the hashtags used, or probable keywords that are related to the domain. This tends to bring a mix of data that is basically dominated by related content. However, it still includes noise in the form of non-relevant data. The first part of the data tidying process involves filtering the retained data using probable keywords for the disaster domain under consideration. Following

this, the tweets generated by newsfeeds are discretized to avoid spamming or non-individual posts. In the next part of data tidying, each tweet content is labeled as relevant, partially relevant or non-relevant, in order to explore content commonalities and differences in the subsequent exploration part. Manual labeling is determined as follows:

- Content related directly to a disaster event is marked as relevant
- Content related to a disaster in a general manner such as criticism of a political party or memory of an old disaster event is marked as partially relevant
- Content related to a very different context to a disaster event such as `I am like a bomb today' or `I am bored to death' is marked as non-relevant.

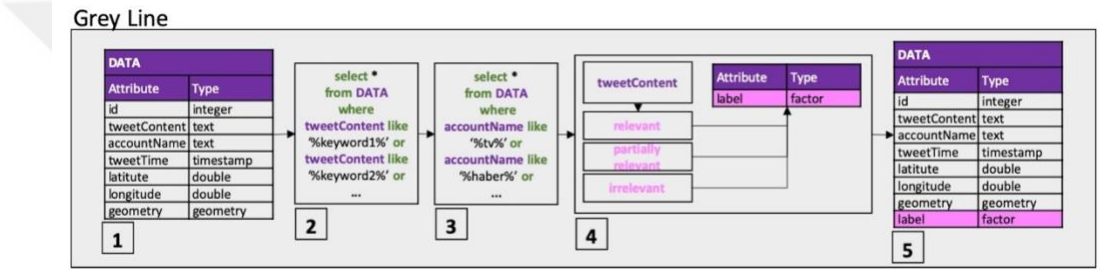


Figure 3.2 : Grey line for data tidying.

The grey line in Figure 3.1 is extended in Figure 3.2 to provide steps of the data flow for data tidying. Each step in Figure 3.2 represents:

1. Acquired data table
2. Pre-filtering using domain-based keywords (rough filtering)
3. Newsfeed accounts removal from the data using media related keywords
4. Manual labeling as relevant, partially relevant, or non-relevant to the domain
5. Label appended to the data table

This process provides pre-filtered and labeled data for the data exploration part. In this data tidying process, there is also a text-cleaning step, which is re-designed as a result of the outcomes of the data exploration. The tr_text_clean (Gulnerman, 2020) function is combined with the following steps:

1. Fix of encoding problems due to different Turkish language characters
2. Removal of emoticons, punctuation, whitespaces, numbers, Turkish and English stop words, long and short strings, and URLs

   This eases the tokenization of each word as part of the next processes with regard to text mining.

3.3.3 Data exploration

Explorative data mining methods are used to identify similarities and differences between relevant/partially relevant and non-relevant chunks of manually labeled tweets. Commonality and comparison clouds, a pyramid plot, and word networks, are visualized to gain insight into the data for further processing.

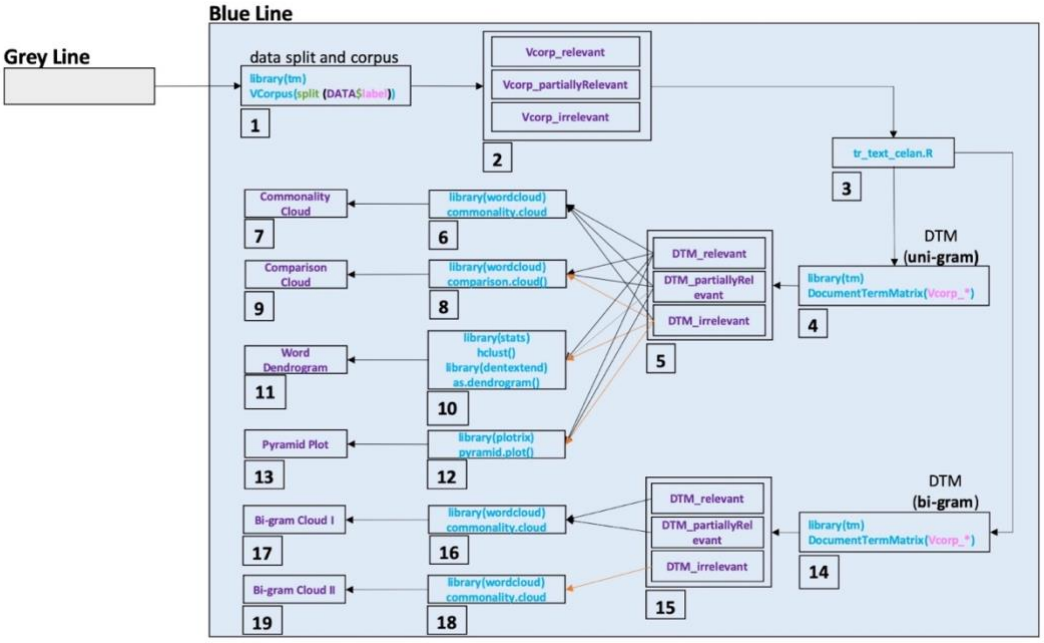


Figure 3.3 : Blue line for data exploration.

The blue line in Figure 3.1 is extended in Figure 3.3 to provide the data flow steps for data exploration purposes. Each step in Figure 3.3 represents:

1. Data division in terms of label type and conversion of data type from data frame to corpus
2. Divided data assignment as corpus in terms of label type
3. Text cleaning for each corpus using the tr_text_clean function
4. Uni-gram Document Term Matrix (DTM) creation for each corpus (each term has a word)
5. Assignment of three DTMs for relevant, partially relevant and non-relevant data
6. Top 100 most frequent words determination in three DTMs
7. Commonality cloud plot with varying word size in terms of word frequency
8. Fifty most frequent distinct terms identification in relevant + partially relevant, and non-relevant DTMs
9. Comparison cloud plot with the determined distinct terms in DTMs
10. Word network based on hierarchical clustering
11. Word dendrogram's plot to reveal the term associations

12. Frequency percentage of commonly used terms in terms of relevant + partially relevant and non-relevant DTMs
13. Pyramid plots with common terms in DTMs ordered by the frequency percentage difference
14. Bi-gram Document Term Matrix (DTM) creation for each corpus (each term has two words)
15. Assignment of three bi-gram DTMs for relevant, partially relevant and non-relevant data
16. Top 100 most frequent bi-gram terms determination in relevant and partially relevant DTMs
17. Commonality cloud plot with varying plot size of bi-grams in terms of frequencies in DTMs
18. Top 100 most frequent bi-grams determination in non-relevant DTM
19. Commonality cloud plot with varying plot size of bi-grams in terms of frequencies in DTM

These help to explore relevant, partially relevant and non-relevant data contents. The working details of the packages used in this data exploration process are given in the following subsections. During this exploration, the importance of suffixes for discrimination in terms of relevancy, and the word association differences in terms of relevancy was explored. In respect to this, word stemming was not applied in the text cleaning function, and the deficit in terms of the use of unigram pre-filtering was revealed.

3.3.3.1 Commonality cloud

This function is deployed as part of the `wordcloud' package (Fellows et al, 2018) in R, runs over a term's frequency, and plots the most frequent `n' number of terms (words) as determined in the function argument. For this study, the top 100 terms are plotted to show which terms have dominance within the dataset.

3.3.3.2 Comparison cloud

As with the commonality cloud, this function is also deployed as part of the `wordcloud' package (Fellows et al, 2018). For a comparison cloud, two chunks of data are required in order to be able to compare the most frequently used terms. In this work, it is used to see which common words are most frequent in both labelled datasets. This function provides insights for sentiment analysis that depends on constituent words scoring without any weights.

3.3.3.3 Pyramid plot

This function is deployed within the `plotrix' package (Lemon et al, 2020) and is used to show frequency differences in both chunks. Frequency difference is normalized with the division of the maximum term frequency in datasets, and ordered in terms of the differences. In this study, the use of a pyramid plot visualizes 50 words in the datasets that have the highest difference rates from each other. This plot is utilized to express the weights of terms that could be used for the intended classifier.

3.3.3.4 Word dendrogram

The `dist' and `hclust' function from the `stats' package (R Core Team, 2013) is utilized to create a word network as a hierarchical cluster, and the `dendextend' (Galili, 2015) package is adopted to visualize and highlight terms in a dendrogram. The Euclidean method is used in the arguments of the `dist' function to determine the distance between terms before the hierarchical cluster process. The dendrograms of relevant, partially relevant and non-relevant sets of data are expressed to determine word association differences.

3.3.4 Data processing

Considering the insight gained from the data exploration part of the methodology, the processing part is carried out using the different filtering techniques based on sentiment lexicons and machine learning. In the first three techniques, the current generic subjective lexicons for the Turkish language (Dehkharghani et al, 2016; Ozturk and Ayvaz, 2018) are exploited to classify a roughly-filtered dataset as being relevant, partially relevant or non-relevant. The fourth and fifth techniques use the manually labelled data to allow building a machine learning classifier for fine-grained filters.

A domain-based perspective is built by exploiting previous similar cases. The use of subjectivity lexicons to filter out non-relevant content and for proposing a way of dealing with terror domain-based data with regard to the Turkish language is a brand-new approach in terms of SMD analysis. Subjectivity lexicons are used to avoid any misunderstandings with regard to Turkish words. For example, the word “bomb” can be used to describe the detonation of an explosive instrument or as the name of a popular dessert in Turkey. This is why valuable information from the posts was not ignored; only relevant classifications were targeted by using the subjectivity lexicons

to pick out the relevant homonymous words in the Turkish language. Yet it clarifies the benefits of the use of word suffixes (i.e. without stemming a word) in Turkish.

From this first perspective, all relevant content is supposed to include negative sentiments, while the non-relevant contents are considered to be positive. The first generic lexicon comprehensively developed for Turkish is STN (Dehkharghani et al, 2016) which has nearly 15,000 terms (uni /bi-gram). STN provides the term's sentiment scores (from 0 to 1) for each sentiment label (positive, negative and objective), and the winning sentiment label for each term. In this study, the first technique scans the constituent words of tweets regarding the sentiment labels for each term, and the content of each tweet is classified in terms of the highest count of the sentiment labels.

As far as the second technique is concerned, the scores of each sentiment label as derived from STN are taken into consideration, and the scanned terms for each tweet are summed to find out the most weighted sentiment relating to that content. The highest scored label classifies the tweet content as being relevant for the highest negative, partially relevant for the highest objective and non-relevant for the highest positive.

The third technique uses another lexicon for the Turkish language (LOA) created by (Ozturk and Ayvaz, 2018) which depends on over 5,000 commonly-used daily words scored from -5 to +5 by three people. The average of these three is accepted as the polarity score for each term. For the third technique, tweets are scanned to match the terms in the LOA, and the score of the matched words are summed to classify the contents, with a minus score identified as being relevant, a zero score as being partially relevant, and a plus score as being non-relevant.

There are few studies on comparisons between text filtering on SMD (Dong and Han, 2004; Hassan et al, 2011; Healy et al, 2005; Trivedi et al, 2015) because most of the machine learning techniques are not considered on SMD. As can be seen from those studies, the most common and efficient techniques that can be used with regard to text filtering are Naive Bayes, Neural Network and Support Vector Machine (Aramaki, 2011; Go et al, 2009; Ikonomakis et al, 2005; Sriram et al, 2010). That is why these three techniques are considered in this study to filter domain-based SMD.

For the fourth technique, a machine-learning perspective is used to create a classifier. The Naïve Bayes (NB) classifier (Wu et al, 2008) is utilized, based on the probabilistic determination method. According to this method, a classifier considers the likelihood probability of each class over a trained dataset. It then calculates the conditional probability of each term seen in the trained dataset regarding each class. The classifier performs this analysis over the test data by multiplying the probability for each term scanned for each tweet, as well as multiplying the likelihood probability of each class. Finally, the classifier compares the probabilities of each class and picks the highest one to label. The method considers the rate of incidence for the commonly-used words in each class. This is advantageous in terms of classification performance with regard to roughly filtered data.

In the fifth technique, the Neural Network (NN) classifiers are trained. The 'nnet' package (Ripley et al, 2016) is utilized to model the classifiers. In terms of the NN principles, for training a dataset, a document term matrix (DTM) is accepted as input signals with class labels. Three classifiers are trained using a multi-class dataset. The weights of hidden layers are not initiated up front, and training is performed with 500 iterations. Considering text classification with DTM is not a linearly-separable problem, and it does need hidden layers when classifying with an NN classifier. Therefore, the hidden layer parameter in the classifier is specified as 1, 2, and 3 to experimentally determine the best hidden layer number. Increasing the hidden layer number provides high classification accuracy, but also creates a time cost. However, the main aim of the study is not finding the best fit for a NN, but assessing ways to determine the outcome of faster filtering techniques when mapping. In respect to this, the NN classifier that shows a reliable statistical performance with 2 hidden layers is accepted in this NN processing part.

In the data processing part, mainly sentiment lexicons and machine learning techniques are used to filter domain-based data. The sentiment lexicons used are the ones for use with the Turkish language. The machine learning techniques used are the ones commonly used for text classification (Aramaki, 2011; Go et al, 2009; Ikonomakis et al, 2005; Sriram et al, 2010). In addition to the NB and NN, Support Vector Machine (SVM) is another popular technique that is used for text classification. As the sixth technique, an SVM classifier is trained. The 'e1071' package (Meyer et al, 2018) which is a misc. functions of the Department of Statistics, TU Wien, is utilized to implement

the classifier. The package allows the training of an SVM classifier by vectorising each term count in each document with the training labels. Since SVM provides successful results on linearly-separable problems, different kernels such as polynomial, radial, and sigmoid are used for the SVM classifiers. Although the SVM classifiers achieve a high degree of accuracy due to the use of unbalanced datasets in this study, they achieve zero sensitive results for two of the three classes. Therefore, it was decided that SVM was not suitable for further assessment in this study.

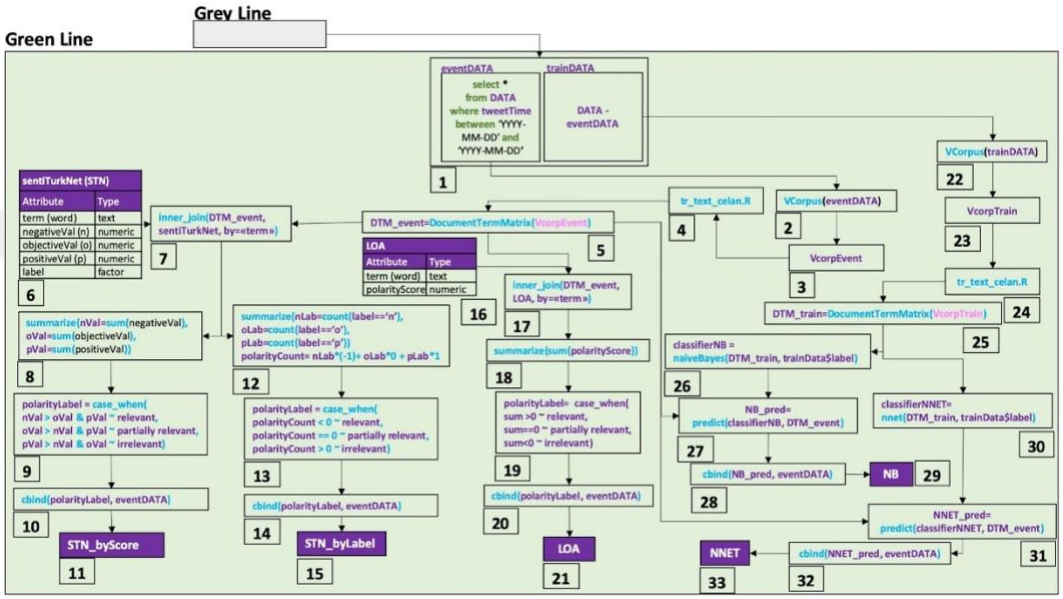


Figure 3.4 : Green line for data processing.

The green line in Figure 3.1 is extended in Figure 3.4 to provide the data flow steps for data processing. Each step in Figure 3.4 represents:

1. Data division as event day data and training data
2. Conversion of data type from data frame to corpus for event day data
3. Data assignment as corpus for event data
4. Text cleaning for event corpus using the tr_text_clean function
5. Uni-gram Document Term Matrix (DTM) creation for the cleaned corpus
6. sentiTurkNet polarity lexicon data frame
7. Inner joining terms in event data DTM and terms in sentiTurkNet
8. Aggregation of negative values, objective values, positive values for terms in each document
9. Polarity decision for each document depending on the bigger values of each aggregation
10. Polarity label appended to event data
11. Assignment of event data with the polarity label as STN_byScore data
12. Count of negative (-1), objective (0), and positive (1) labels of terms in each document

13. Polarity decision for each document depending on the bigger count
14. Polarity label appended to event data
15. Assignment of event data with the polarity label as STN_byLabel data
16. Inner joining terms in event data DTM and terms in LOA
17. Aggregation of polarity score of terms in each document
18. Determination of polarity label based on sign (-, +) of polarity score aggregation
19. Polarity label appended to event data
20. Assignment of event data with the polarity label as LOA data
21. Conversion of data type from data frame to corpus for training data
22. Data assignment as corpus for training data
23. Text cleaning for training corpus with tr_text_clean function
24. Uni-gram Document Term Matrix (DTM) creation for the cleaned training corpus
25. Modelling a naïve Bayes classifier with the training DTM
26. Prediction of event DTM relevance with the naïve Bayes classifier
27. Prediction label appended to event data
28. Assignment of event data with the prediction label as NB_pred data
29. Modelling a neural network classifier with the training DTM
30. Prediction of event DTM relevance with the neural network classifier
31. Prediction label appended to event data
32. Assignment of event data with the prediction label as NNET data

This provides us with classified event data as a result of the five different techniques. The classified data in terms of different techniques is further processed with the methodology given in subtitle 3.3.5 for quantifying the impact of filtering on spatial accuracy.

If the language is different than the language of this study, the users should alter several steps that were shown within the methodology. In the data tidying step, data should be fixed due to encoding problems caused by different Turkish language characters. This tidying operation is specific to the Turkish language and should be changed for other languages that have different encodings. In addition to that, the stop words removal of the text cleaning part should be changed into the defined language if the language is different than the Turkish. Another alteration in the tidying part might be the addition of the “stemming” process. This study omitted the “stemming” since the suffixes contribute to the discrimination of the relevance as it is discovered in the data exploration part. However, the “stemming” operation could contribute to the discovery of the very related content in other languages. In the data processing part, the sentiment

lexicons are the ones for Turkish Language and these are also required to be changed if the filtering process will be carried out with the sentiment lexicons of a different language. The domain in this study is defined as the terror attacks and this should certainly contain negative sentiment. As it was also discovered in the data exploration part, irrelevant data contain positive sentiments. That is why the use of sentiment lexicons can be successful for this terror domain. That means sentiment lexicons use should be avoided for determining the relevancy of such domains that can contain all types of sentiments mostly the positive ones. For instance, in a marketing research, the domain may include both positive, neutral or negative sentiments, therefore the use of sentiment lexicons is not plausible for filtering relevant data.

3.3.5 Interpretation of spatial SMD

Accurate text mining for fine filtering is the first aspect associated with producing more reliable maps for domain-based SMD analyses. Yet on its own, it is not enough to determine the spatial accuracy of the maps produced from the SMD. To determine how different filtering techniques impact on spatial accuracy, this study proposes a spatial interpretation method. The interpretation method involves two main steps - spatial clustering and spatial similarity calculation. In subsection 3.3.6.1, the details of the spatial clustering is given for incidence mapping the filtered data that is used for this study. In subsection 3.3.6.2, the current coefficients for the spatial similarity calculation are introduced. In subsection 3.3.6.2, the new spatial similarity coefficient entitled the Giz Index is proposed with quantitative comparisons of the previously-introduced indices over a designed test data. The similarity scores vary in terms of the chosen similarity coefficients. Therefore, it is important to determine which spatial similarity index is used. In subsection 3.3.6.2, a test designed for the comparisons of similarity indices is used to reveal which index returns plausible similarity value for incidence map comparisons.

3.3.6 Spatial clustering

There are a number of spatial clustering algorithms (Han et al, 2001) and methods used for spatial event detection, whose results vary in terms of algorithm selection and methodologies (Ozdikis et al, 2017) Considering the conflicts between different clustering techniques and pre-specified parameters (such as the number of the clusters and required minimum number for each cluster), the Getis-Ord spatial clustering

algorithm (Getis and Ord, 2010; Ord and Getis, 1995) is chosen for each dataset to compare spatial variances due to the different filtering methodologies applied previously. While performing Optimized Hot Spot Analysis located in the Spatial Statistics toolbox in ArcMap (Scott and Janikas, 2009), the cell size is defined as 500 meters in terms of street-level resolution (Middleton et al, 2018) as it is intended to be fine-grained, and the aggregation method is determined as a hexagon polygon to allow us to use the connectivity capacity of a clustering lattice shape (Birch et al, 2007). This first step provides Hot Spot maps for different filtering methodologies, which serves as base maps for the next step - spatial similarity calculation.

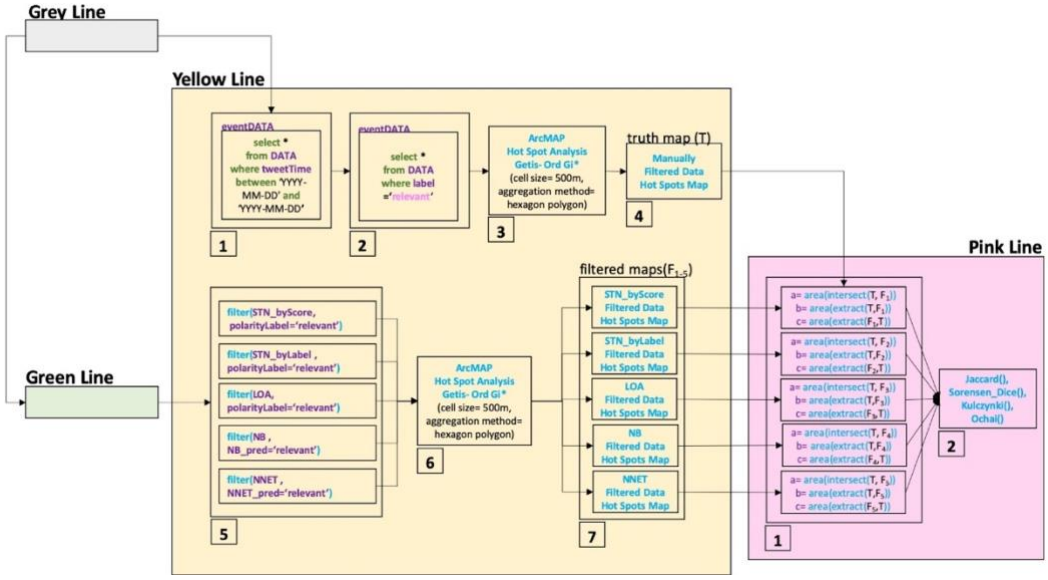


Figure 3.5 : Yellow and pink lines.

The hot spot mapping details of the filtered data are given by the yellow line part in Figure 3.5. Each step in Figure 3.5 represents:

1. Filtering data in terms of day of the event
2. Relevant data selection in terms of manually labeled data
3. Production of a hot spot map for manually filtered relevant data
4. Assignment of the manually filtered data hot spot map as the ground truth map (T)
5. Relevant data selections in terms of each automatically classified data
6. Production of hot spot maps for each automatically filtered relevant data
7. Assignments of the automatically filtered data hot spot maps as filtered maps ($F_{1\text{-}5}$)

This provides ground truth and filtered data hot spot maps for the spatial similarity calculation described in the next subsection.

3.3.6.1 Spatial similarity calculation

In various fields such as biology (da Silva Meyer et al, 2004; HUBÁLEK, 1982), ecology (Michael, 1920), and image retrieval (Smith and Chang, 1996), numerous similarity indices have been proposed. Choi, Cha, and Tappert (2010) present an extensive survey of over 70 similarity measures in terms of positive and negative matches and mismatches when it comes to comparisons. Similarity measures are also in use and vary for the comparison between two maps in order to find semantic similarities between land use, land cover classifications, temporal change and hotspot maps overlaps (Arnesson and Lewenhagen, 2018; Feng and Flewelling, 2004; Hu et al, 2016). In this study, the similarity between incidence maps is tested with the use of similarity coefficients. Incidence mapping similarity is a special case for the similarity measures since the incidences cover a small proportion of the pre-defined area (such as a city, a region or a district). Therefore, the negative matches within the observed area should be disregarded in order to avoid a high degree of misleading similarity due to the high cover of negative matches. (Arnesson and Lewenhagen, 2018) propose the use of four different quantitative similarity measures for determining the similarities or differences between hotspot maps. These are the Jaccard Index (3.1) (Real and Vargas, 1996), the Sorensen-Dice Index (3.2) (Dice, 1945; Sørensen et al, 1948), the Kulczynski Index (3.3) (Kulczyński, 1928) and the Ochai Index (3.4) (da Silva Meyer et al, 2004) each of which are utilized for similarity calculations. These measures return a similarity score of between 0 and 1 regarding True-True (a), True-False (b) and False-True (c) hotspot counts in a comparison between truth and prediction spots. These indices disregard negative matches (False-False (d)) (Arnesson and Lewenhagen, 2018; da Silva Meyer et al, 2004) that are the first requirement for incidence mapping similarity calculations. However, these indices are not specifically designed for spatial purposes, given that they disregard the proximity between mismatching hotspots to the positive matches. In this respect, a new spatial similarity index is formulated for quantifying incidence mapping reliability with regard to the ground truth data.

$$Jaccard\ Index = \frac{a}{a+b+c}, \tag{3.1}$$

$$Sorensen\_Dice\ Index = \frac{2a}{2a+b+c}, \tag{3.2}$$

$$Kulczynski\ Index = \frac{1}{2}\left(\frac{a}{a+b} + \frac{a}{a+c}\right), \tag{3.3}$$

$$Ochai\ Index = \frac{a}{\sqrt{(a+b)(a+c)}}, \quad (3.4)$$

Similarity index calculation details are given in the pink line in Figure 3.5. Each step in the pink line represents:

1. Variable calculations for each comparison (similarity calculation):

- a= area(intersect(T, F_1)) (the area of spatial intersection of the truth map and the filtered map)
- b= area(extract(T,F_1)) (the residual area of the truth map)
- c= area(extract(F_1,T))(the residual area of the filtered map)

2. Calculation of similarity between the truth map and filtered maps with the indices (3.1)(3.2)(3.3)(3.4)

This provide the four spatial reliability scores in terms of current similarity measures. However, the applicability of the measures is uncertain with regard to the incidence mapping since the measures do not consider the proximity of the non-intersection areas, even if they are neighboring. That is why the Giz Index is proposed as a spatial similarity index for incidence mapping studies. In the next subsection, the Giz Index is formulated, and a comparison test between indices is provided over test data. This comparison is provided to explain the eligibility of the Giz Index for incidence mapping comparisons, while the others are not appropriate for every circumstance.

3.3.6.2 Giz index

In a spatial context, omitting negative matches allows a more specified representation of hotspot similarities between two maps. Otherwise the results might be dominated by negative matches and return over 99% similarity for almost all comparisons. Yet the use of coefficients for spatial purposes has some weaknesses such as disregarding the size of separate clusters and the proximity between non-intersecting clusters. The claim here depends on Tobler's First Law of Geography; *'everything is related to everything else, but near things are more related than distant things'*. So, the distance and the size of each cluster needs to be considered in order to draw an accurate spatial similarity index regarding proximate things that might influence each other. A new spatial similarity index has been developed to represent to extent to which two maps are approximately similar to each other. This is entitled the Giz Index. This index is

referred to within the formulas and figures as GI. This index considers each hot spot cluster in the first map as truth ($c_{1..n}$) and in the second map as prediction ($k_{1..n}$).

The methodology of the index detects the area of each cluster ($A_{c1..n,\ k1..n}$), and considers whether they are intersecting to form an area of intersection ($AI_{c1\text{-}k1}$) or non-intersecting to form a residual area of intersection ($ANI_{c1\text{-}k1}$) and the distance between clusters ($D_{c1\text{-}k1}$). The Giz Index is formulated as in (3.5) for single cluster similarity.

$$Giz\ Index = \begin{cases} \frac{AI_{c1-k1}}{A_{c1}} * \frac{\sqrt{A_{c1}}}{\sqrt{A_{c1}}+D_{c1-k1}} + \frac{ANI_{c1-k1}}{A_{c1}} * \frac{\sqrt{A_{c1}}}{\sqrt{A_{c1}}+D_{c1-k1}} & for\ ANI_{c1-k1} < A_{c1} \\ \frac{AI_{c1-k1}}{A_{c1}} * \frac{\sqrt{A_{c1}}}{\sqrt{A_{c1}}+D_{c1-k1}} + \frac{A_{c1}}{ANI_{c1-k1}} * \frac{\sqrt{A_{c1}}}{\sqrt{A_{c1}}+D_{c1-k1}} & for\ ANI_{c1-k1} > A_{c1} \end{cases}, \quad (3.5)$$

In the first part, the equation handles the similarity of intersecting areas (AI). AI similarity is obtained by dividing intersecting areas ($AI_{c1\text{-}k1}$) of the cluster (A_{c1}) and multiplying them with the normalized distance between clusters (ground truth and prediction). This normalized distance is taken as 1 for the intersecting areas part, and is always less than 1 for a non-intersecting area. At the same time, as the distance increases between clusters (c_1 - k_1), the normalized distance will decrease relative to the square root of the cluster area. In the second part, non-intersecting part similarity is calculated with the very same formula as in the first part. There is a condition based on the sizes of the A_{c1} and $ANI_{c1\text{-}k1}$, where the area ratio in the second part of the formula is inversed. For instance, if the area of the first cluster (c_1) is 4 units, and the non-intersecting part is 10 units, the ratio is inversed to become 4/10 instead of 10/4.

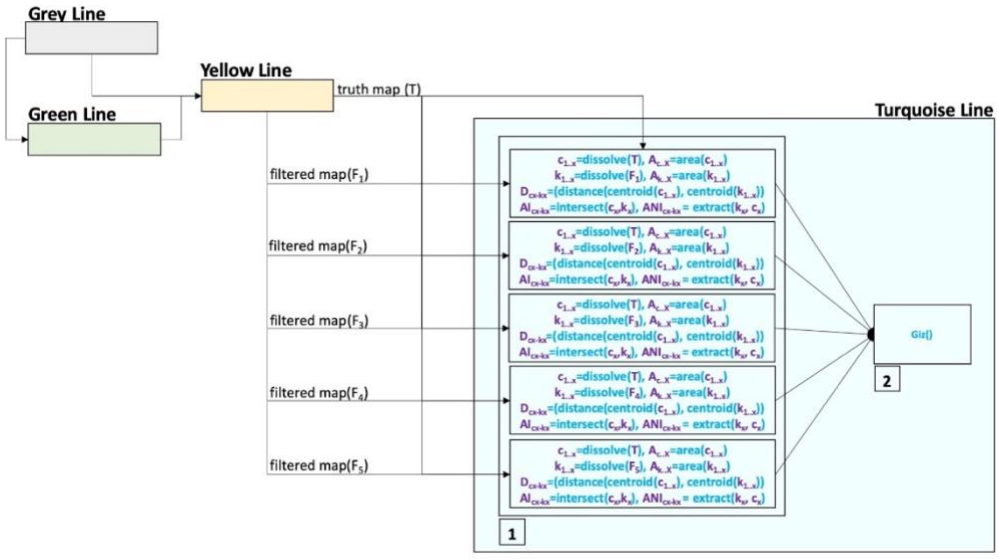


Figure 3.6 : Turquoise line.

The Giz Index calculation details are given by the turquoise line in Figure 3.6. Each step in Figure 3.6 represents:

1. Variables' calculation for each comparison (similarity calculation) for the Giz Index:

- $c_{1..x}$= dissolve(T) (determination of clusters for the truth map by merging adjacent hexagons)
- $A_{c..x}$=area($c_{1..x}$) (defining area of each cluster in the truth map)
- $k_{1..x}$=dissolve(F_x) (determination of clusters for the filtered map by merging adjacent hexagons)
- $A_{k..x}$=area($k_{1..x}$) (defining area of each cluster in the filtered map)
- $D_{cx\text{-}kx}$ = (distance(centroid($c_{1..x}$),(centroid($k_{1..x}$))) (defining distance between each truth map cluster and each filtered map cluster) (This step is important in order to define which cluster in the filtered map is compared to which truth map cluster. The closest distance determines the candidate cluster in the truth map.)
- $AI_{cx\text{-}kx}$ = intersect(c_x, k_x) (defining the area of the intersecting clusters)
- $ANI_{cx\text{-}kx}$ = extract(k_x, c_x) (defining the area of non-intersecting cluster parts)

2. Calculation of similarity between truth maps and filtered maps using the Giz Index (5)

This provides similarity measures weighting the degree of similarity depending on the proximity between truth map clusters and filtered map clusters.

The Giz Index is tested and compared with the current indices with regard to the test data that is displayed in Figure 3.7. The test data includes six instances and each instance has two predictions (p1, p2) and one truth cluster. The test data is developed to reveal the response details of the similarity indices in terms of the different sizes of prediction clusters and the distances between the prediction clusters to the truth cluster.

The similarity between the prediction clusters (p1, p2) and the truth cluster for each instance is calculated using the current indices and the proposed Giz Index. The similarity results vary between 0 and 1 and are presented in Table 3.1. In the first three instances, there are no intersecting clusters. Therefore, the results of the current indices return 0 value either the prediction is proximate or far to the truth cluster. Also, these indices disregard the non-intersecting cluster size in these zero intersection circumstances. On the other hand, the Giz Index returns varying similarity values in terms of the proximity and size of the non-intersecting clusters - in these instances, I, II and III. When differences in size and distance are higher, the Giz Index (GI) is closer to 0, otherwise, it is closer to 1. The scores of the current similarity indices respond to

the size of the intersecting area in instance IV, V, and VI. When the intersection is proportionally higher, the current scores become higher as in the GI. However, the residual area of the cluster has zero addition to the similarity increase even if they are placed in neighboring parts of the truth cluster. In respect to the results, it is obvious that the current similarity indices totally disregard non-intersecting parts, when there is no intersecting cluster. Therefore, the current indices might cause misinterpretations in incidence mapping studies since the near incidences to the real incidence areas have value when it comes to determining the exact incidence area. In other words, the occurrence of close incidences on a map can be interpreted as signs indicating the main event. Therefore, the proximity of non-intersecting clusters and size should be considered for the correct interpretation of the maps.

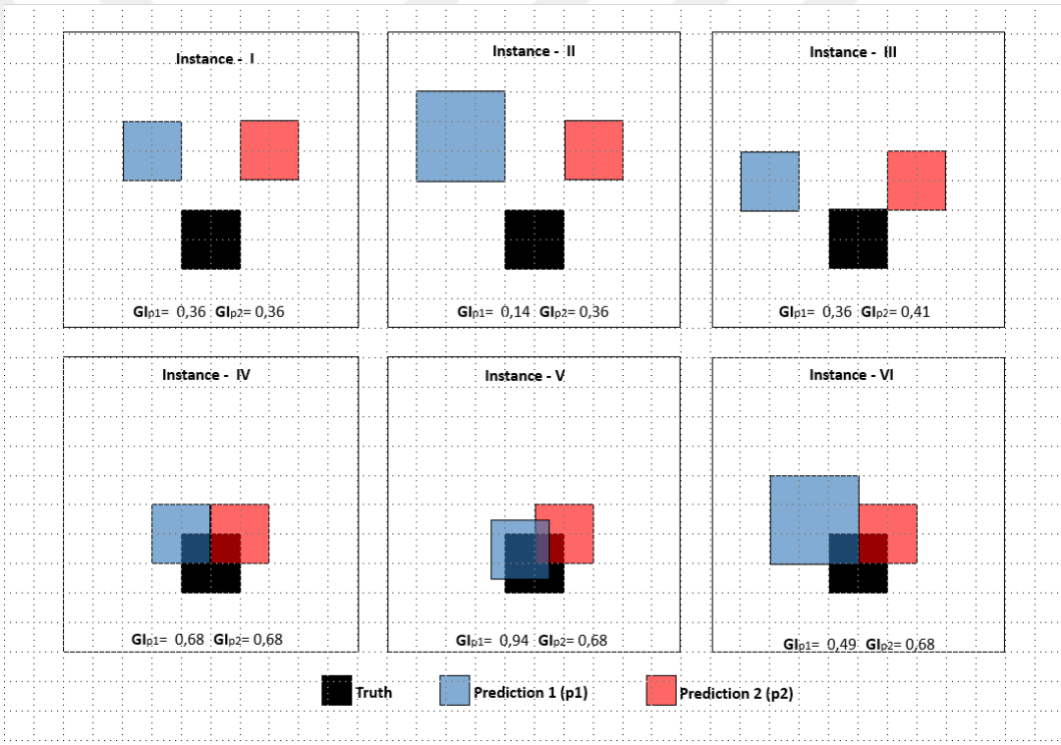


Figure 3.7 : GI similarity test over 6 instances.

Table 3.1 : Similarity scores of test data over six instances in terms of similarity indices.

	Instance - I		Instance - II		Instance - III		Instance - IV		Instance - V		Instance - VI	
	p1	p2	p1	p2	p1	p2	p1	p2	p1	p2	p1	p2
JI	0	0	0	0	0	0	0,14	0,14	0,39	0,14	0,08	0,14
SI	0	0	0	0	0	0	0,25	0,25	0,56	0,25	0,15	0,25
KI	0	0	0	0	0	0	0,25	0,25	0,56	0,25	0,18	0,25
OI	0	0	0	0	0	0	0,25	0,25	0,56	0,25	0,17	0,25
GI	0,36	0,36	0,14	0,36	0,36	0,41	0,68	0,68	0,94	0,68	0,49	0,68

The Giz Index supports exploratory inferences when the quantitative measure is between 0 and 1, by considering both the size of spatial clusters and their proximity. The size and location accuracy are important for the several subjects that form the domains, and the Giz Index can provide quick and automatic spatial interpretation of the analyzed data. This Index can also be used for a quick look at the validation processes with regard to social media data if there are any other secondary data sources that can be accepted as the truth. The Giz Index can also be used for various domains such as concerts, elections, marketing in addition to disasters as are mentioned in this study. The spatial similarity introduced in this research can be used for the spatial comparison of simulation maps and estimation maps, and the resulting maps associated with different studies using different methodologies can be used for similar purposes.

3.4 Case Study

3.4.1 Importing and retaining data

In this study, 8 months of data from May to December 2016 was selected regarding 10 terrorist attacks that occurred in Turkey. During that period, geo-referenced tweets were captured and inserted in the PostgreSQL database using GTD.

3.4.2 Data tidying

Pre-filtering was done roughly using keywords such as `attack' (saldırı in Turkish), `bomb' (bomba in Turkish) and `explosion' (patlama in Turkish) to provide targeted chunks of data with a mix of relevant and non-relevant content. While filtering, all combinations of case-sensitive and Turkish character encodings were considered in order to retrieve all possible related content such as `saldırı', `SALDIRI', `Saldiri' etc.

Following this, 285 tweets from 6 newsfeed accounts were detected and removed from the pre-filtered data due to possible content being mixed with diverse and non-relevant news. After removal, a total of 4,395 tweets were classified manually using three labels; 1- relevant (RL) - pertaining to a newly-occurred terror attacks, 2- partially relevant (PR) - related content including terror in general such as political party criticism or memories of an old terror event, 3- non-relevant (IR) -- contents totally unrelated to a terror attack such as `I love bomb dessert' or `Go team go, attack and beat them'. As labelled, 934 of the tweets were non-relevant, 799 were partially

relevant, and 2,662 tweets were directly relevant to a terror event. Maps of each set of IR, PR, and RL tweets were created and classified by month (Figure 3.8).

One of the inferences while labelling is that social media users react to newly-occurred terror attacks by protesting about terrorists, the security services, the government and political parties. However, many of the tweets include condolences to the families of the victims regarding second-hand information obtained from somewhere else such as traditional/news media. Very few of the users were sharing information about witnessing the incidents. Consequently, these inferences predominantly indicate a repetition of similar content. Thus, it can be said that mining for first-hand information is just like looking for a needle in a haystack. This first-hand information is of great importance for incidence mapping during or shortly after a disaster, as implied in the motivation for this study.

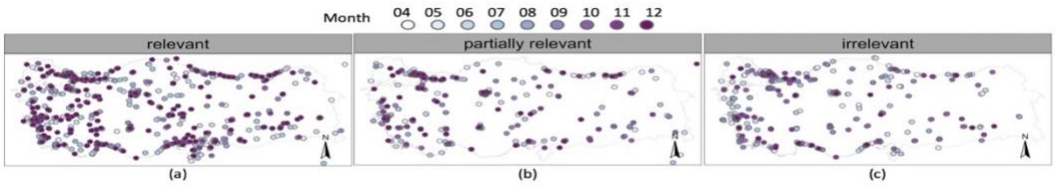


Figure 3.8 : Labelled pre-filtered spatial data (**a)** RL; **(b)** PR.

Prior to explorative data analysis, the text of the tweets was cleaned to avoid later problems when it came to processing a better bag of words. Therefore, emoticons, punctuation, stop words in Turkish (Aksoy and Ozturk, 2018) and English, numbers, and URLs were removed from the tweets, and all encoding and letter case problems were fixed. In addition, city and county names with regard to Turkey that may create a bias in terms of this work, and were not needed while classifying content, were removed from the texts. Stemming, which is one of the further text cleaning steps, was ignored in this study. In this way, the aim was to preserve rich meaning differences with suffixes in Turkish. By using the tr_text_clean function, text cleaning of this work was undertaken on the one hand by including several functions from the TM package (Ingo Feinerer et al, 2013) and the ggrepel package (Slowikowski, 2019) and, on the other, self-defined sub-functions. Details of the tr_text_clean cleaning function can be found in (Gulnerman, 2020) in terms of its use in similar works.

3.4.3 Data exploration

Data exploration for Turkish terror domain-based data was handled descriptively using commonality and comparison clouds, a pyramid plot and a word dendrogram in order to show differences between terms (words) with regard to data relevant to the domain.

(a) **(b)** **(c)** **(d)**

Figure 3.9 : Keyword filtered tweets **(a)** commonality cloud; **(b)** comparison cloud; **(c)** bi-gram word cloud for relevant tweets; **(d)** bi-gram word cloud for non-relevant tweets.

Commonality for uni/ bi-grams and comparison clouds are visualized (Figure 3.9) after the several cleaning functions are applied to all three chunks of data. The commonality cloud plotted for uni-gram (a) represents the most frequent terms over all the labelled chunks, in which the relevant chunk dominates the frequency due to the relatively high number of tweets compared with other chunks. This means that hot topics dealt with during the disaster could be easily estimated by the use of the commonality cloud over the roughly-filtered data. The comparison cloud (b) displays the most frequent terms in both chunks, in terms of relevant/ partially relevant tweets and non-relevant tweets. The comparison cloud provides a more detailed explorative look by plotting frequent terms regarding non-relevant and relevant (partially or totally) chunks of data (b). Terms in this study are in Turkish and translated into English, with the words in parenthesis being the actual Turkish words encountered in this study. The first inference is that chunks include common terms such as `bomb' (bomba), `explosion' (patlama) and `attack' (saldırı); dissimilar frequent terms include words such as `dessert' (tatlı), `energetic' (enerji), `mercy' (rahmet)). The second inference is that some of the most common words in the comparison cloud have taken different suffixes. For instance, `attack' in the relevant chunk has noun suffixes, while in the non-relevant chunk it is mostly used as a verb in the imperative mood. Therefore, these were assessed as different terms by omitting stemming in the pre-processing period.

In addition to the unigram word cloud, bi-gram word clouds are plotted for both chunks, whether the use of any bi-gram could be pervasive in either one or both (Figure 3.9 (c), (d)). For instance, `God's mercy' (allahtan rahmet) is the most frequent bi-gram in the relevant (partially or totally) chunk, while `coming like a bomb' (bomba geliyoruz) is the most frequent bi-gram in the non-relevant chunk. From the perspective of the subjective lexicons, we might consider that `God' can be labelled as positive. However, in conjunction with `mercy' it turns into a condolence phrase that has a negative sentiment. In the same vein, `bomb' can be perceived as being negative or positive, while on the other hand `coming like' expresses a positive sentiment, with `attack' indicating negative sentiments. Those words are not `negators' that directly invert a polarized meaning like the `not' effect over adjectives, as in `good' and `not good'. It shows the importance of word association in order to identify the correct sentiment.

A word dendrogram for each chunk is visualized to compare the changing word association between chunks. There are a couple of inferences from the dendrograms. The first one is that there are few branches for the same words with different suffixes in the same dendrogram, and also in different ones (such as; terör, terörü, teröre and terörsüz etc. colored as red in Figure 3.10). These suffixes provide a clue about the other parts of the sentence. For example, `terror' (teröre) could be completed with `damn' (lanet) to create `damn terror'. Similarly, while the word `terörsüz' means `without terror', it could be completed with terms such as `wishing for a day without terror' (colored as blue in Figure 3.10). In addition, the word for `explosion' (patlama) in a relevant dataset, is not directly associated with any word, while in the non-relevant chunk it is associated with `good morning' (günaydın), `energy' (enerji) and sampled as `Good morning! I have had an energy explosion today' (colored as orange in Figure 3.10).

Furthermore, the leaves of dendrograms display commonly-associated words such as `wish' `speedy' `recovery' `to the wounded' (`diliyorum', `acil', `şifalar', `yaralılara') in a relevant dendrogram, `counter' (`mücadele' in Turkish), `terrorism' (`terörle'), `branch' (`şube'), `directory' (`müdürlüğü') in partially relevant tweets, and `ready' (`hazır'), `to explode' (`patlamaya') in non-relevant sets (colored as green in Figure 3.10). Therefore, keeping suffixes is important, as well as assessing word associations in the same document to determine the exact sentiment.

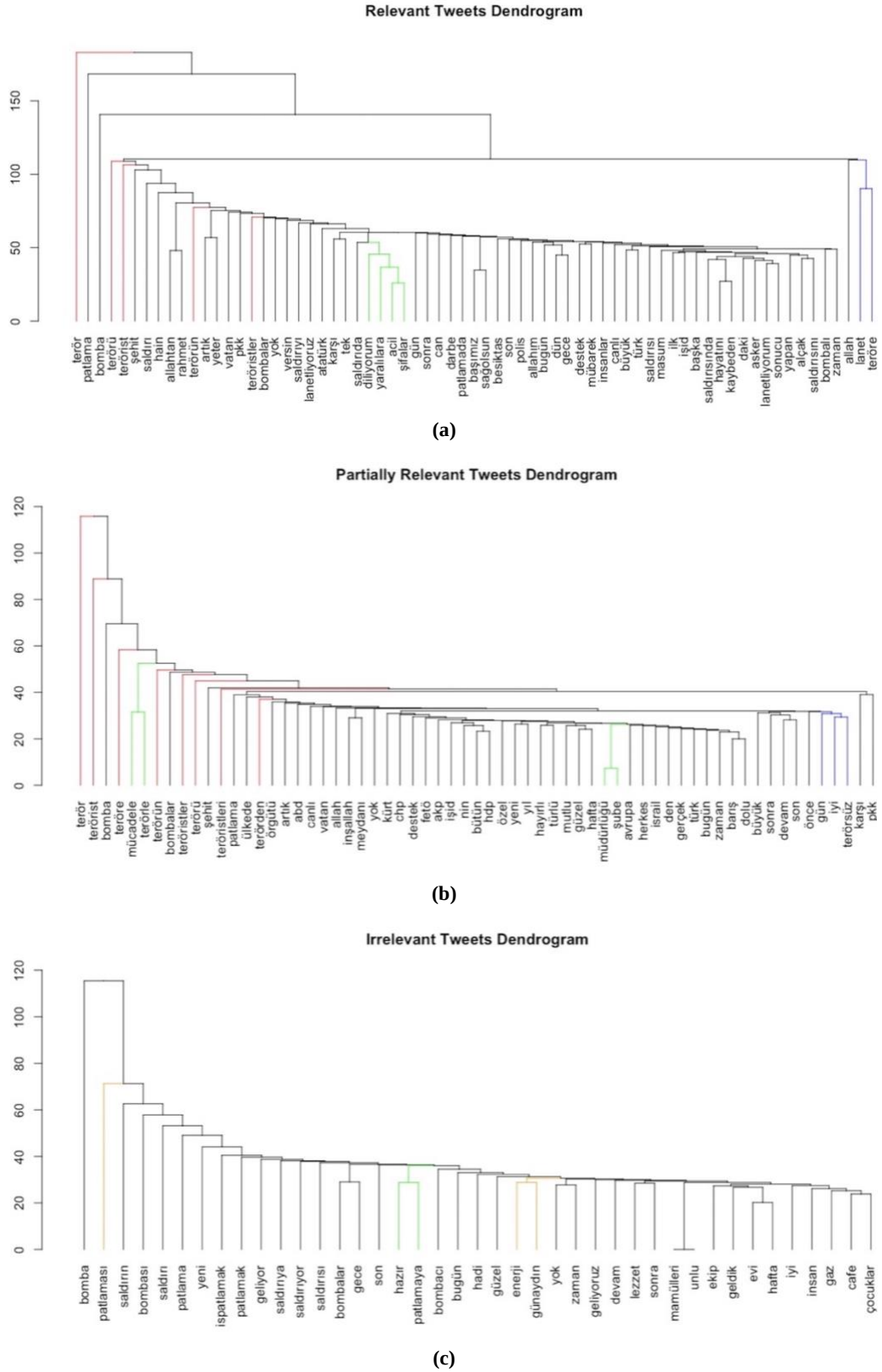


Figure 3.10 : Word Association Dendrograms for **(a)** RL; **(b)** PR; **(c)** IR tweets.

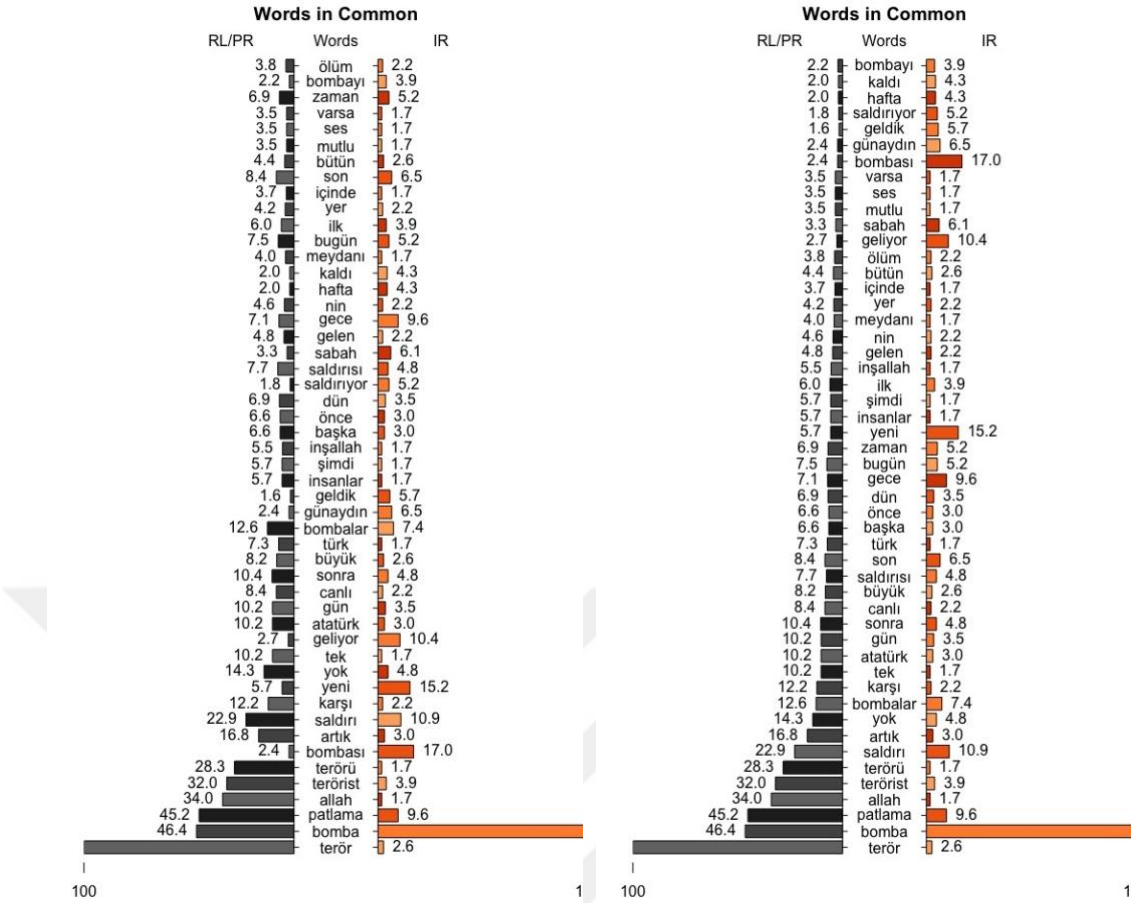


Figure 3.11 : Pyramid plot **(a)** ordered by the term frequency difference between RL/PR and IR; **(b)** ordered by the term frequency in RL/PR.

The use of a pyramid plot also offers another chance to explore the dis/similarity in terms of frequency between RL-PR and IR. The terms (`terör' in RL-PR, `bomba' in IR) which have the maximum frequencies in both chunks, are assumed to be 100 units, and the frequency values of the other terms are normalized using a rate ratio (3.6). Normalization is applied to RL/ PR and IR chunks separately in order to avoid the dominance of different chunks sizes.

$$ntfv_i = \frac{tf_i * 100}{\max(tf)}, \tag{3.6}$$

The pyramid plot used in this study plots 50 common terms that have the highest frequency difference between the RL/PR and IR chunks ordered by the normalized term frequency value difference (ntfv) (Figure 3.11 (a)) and ordered by the ntfv of the RL/PR chunk (Figure 3.11 (b)). This plot is important in that it shows that although there are common words, each has a different frequency rate in the various chunks.

For instance, keywords such as `terör', `bomba', `patlama', used for rough filtering, are unsurprisingly placed at the bottom of the plot as the highest ntfv for the RL/ PR chunk, while they have 2.6, 100, 9.6 ntfv respectively for the IR chunk. In addition, the plot provides the feasibility of looking more closely at the ntfv with regard to the same words with different suffixes as in `terör', `terörist', `terörü' and `bomba', `bombalar', `bombası', `bombayı'.

3.4.4 Processing data

The explored data is processed using two lexicon libraries (STN and LOA) with three different processes (score, score label and polarity labels) in order to determine the best fine-grained filtering process and its performance. The general, the terror domain filtering capability of current Turkish subjectivity lexicons are revealed before the implementation of the techniques with regard to two cases from the terror domain. As a result, all self-labelled data are processed with the current subjectivity lexicons in the form of STN and LOA. This resulted in the prediction of each tweet's sentiments by separately utilizing terms with scores in STN, polarity labels in STN and scores in LOA. To determine the content of each tweet, whether negative, positive or objective, each tweet's constituent terms' label counts and score sums were calculated. The relevant metric for this score was chosen in terms of its overall accuracy (3.7) which is the ratio of the correctly classified cases vs. the total data count.

$$Overall\ Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Data}, \tag{3.7}$$

These three processes respectively indicate that 36%, 40% and 49% of the results are accurate for filtering. However, the first two STN processes resulted in 37% being non-applicable (NA), and the last process (LOA) resulted in 5% being non-applicable (NA) (Table 3.2 (a), (b), (c)). Thus, a 37% NA result can be interpreted as being inadequate in terms of the STN for the terror domain, while LOA has a wider capability to cover terms with just 5% NA results. This might indicate that LOA has more words that intersect with the domain terms. On the other hand, the overall accuracy for both lexicons is not sufficiently adapted to that domain to allow fine-grained filtering.

Table 3.2 : Confusion matrices of sentiment analysis for fine-grained filtering by (a) STN by score; (b) STN by polarity label; (c) LOA by score.

(a) STN by score		Truth		
		R	IR	PR
Prediction	R	585	88	87
	IR	294	130	131
	PR	734	440	263
	NA	1049	276	318
		Overall Accuracy 36%		

(b) STN by polarity label		Truth		
		R	IR	PR
Prediction	R	991	549	382
	IR	617	106	97
	PR	5	3	2
	NA	1049	276	318
		Overall Accuracy 40%		

(c) LOA by score		Truth		
		R	IR	PR
Prediction	R	1564	435	465
	IR	402	169	143
	PR	316	160	129
	NA	0	170	62
		Overall Accuracy 49%		

R (positive), IR (negative), PR (objective)

The explored dataset includes tweets pertaining to several terror attacks that have occurred in Turkey. Two cases are picked for a comparison of filtering techniques, and in terms of mapping reliability. In the first case, the test data is roughly-filtered data that was generated after the terror attack which occurred near the Besiktas Vodafone Arena (BVA) in the Besiktas district of Istanbul. The second case data is roughly filtered after the Ataturk airport terror attack (ATA). Since these two terror attacks caused many casualties and fatalities, and created a threat to thousands of people in the urban area. Many people expressed their sorrow, offered condolences, and express their hatred of the attacks. Therefore, sentiment analysis is seen as a way of undertaking fine-grained filtering. Both datasets have been processed with regard to the three lexicons and two machine learning-based techniques. The first three techniques were applied regarding each case of datasets. However, these processes returned some unlabeled outputs (NA), which are disregarded for the calculation in terms of accuracy. The confusion matrices of the first set of techniques are displayed in Table 3.3 (a)(b)(c) for the BVA and Table 3.4 (a)(b)(c) for the ATA. With regard to the fourth process, a Naïve Bayes classifier was trained and tested for the same events. The rest of the roughly-filtered and labelled data was used as a training dataset to build a Naïve Bayes classifier. As a result of this process, the classifier labelled all data with

87% accuracy (Table 3.3 (d)). This approach was tested with the second dataset generated after the ATA. The ATA event resulted in 84% accuracy (Table 3.3 (d)) using the same approach. The fifth process was a neural network (NN) classifier, which was trained using the manually labelled data. The model classified BVA and ATA data with 61% and 70% accuracy (Table 3.4 (d) (e)) respectively. A hidden layer parameter was assigned as 1, 2, and 3 in order to find the best model for NN. In the NN processing part, an NN structure consisting of 2 hidden layers is adopted due to its reliable statistical performance.

A confusion matrix is plotted for the evaluation of each filtering technique with overall accuracy, sensitivity (recall), specificity, PosPredValue (precision), F1, F2 and G-Mean (geometric mean) scores. These scores are presenting the details of the filtering performance. While accuracy indicates the general performance of the filtering process, it is not adequate on its own in some circumstances e.g. a misleading high accuracy score can be seen for the unbalanced data classes due to zero sensitivity on all classes other than the major class (Bekkar et al, 2013). Data that is used in this study is unbalanced for the relevant (R) class, since the pre-filtering is applied for the day of the terror attacks. Although the classifiers are modelled to classify multi-classes, it is more important that the major class (R) is classified. In this circumstance, sensitivity, F1 and G-Mean metrics are proposed by the studies dealing with such data and classifiers (Bekkar et al, 2013; Branco et al, 2016; Sun et al, 2009).

Table 3.3 : Confusion matrices of fine-grained filtering processes over the BVA data.

(a) STN by score		Truth		
		R	IR	PR
Prediction	R	203	4	3
	IR	84	0	6
	PR	172	2	5
	NA 44%	358	3	7
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 43%	Sensitivity	0.44	0.00	0.36
	Specificity	0.65	0.81	0.63
	PosPredValue	0.97	0.00	0.03
	F1	0.61	NaN	0.05
	F2	0.50	NaN	0.11
	G-Mean	0.43	0.00	0.44
(b) STN by polarity label		Truth		
		R	IR	PR
Prediction	R	168	4	2
	IR	69	0	5
	PR	222	2	7
	NA 44%	358	3	7
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 37%	Sensitivity	0.37	0.00	0.50
	Specificity	0.70	0.84	0.52
	PosPredValue	0.97	0.00	0.03
	F1	0.53	NaN	0.06
	F2	0.42	NaN	0.12
	G-Mean	0.44	0.00	0.46
(c) LOA by score		Truth		
		R	IR	PR
Prediction	R	429	6	17
	IR	110	0	1
	PR	96	0	2
	NA 22%	182	3	1
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 65%	Sensitivity	0.68	0.00	0.10
	Specificity	0.12	0.83	0.85
	PosPredValue	0.95	0.00	0.02
	F1	0.79	NaN	0.03
	F2	0.72	NaN	0.06
	G-Mean	0.23	0.00	0.29
(d) Naïve Bayes		Truth		
		R	IR	PR
Prediction	R	721	2	12
	IR	57	7	4
	PR	39	0	5
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 87%	Sensitivity	0.88	0.78	0.24
	Specificity	0.53	0.93	0.95
	PosPredValue	0.98	0.10	0.11
	F1	0.93	0.18	0.15
	F2	0.90	0.34	0.20
	G-Mean	0.64	0.85	0.48
(e) Neural Network		Truth		
		R	IR	PR
Prediction	R	506	3	13
	IR	106	3	1
	PR	205	3	7
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 61%	Sensitivity	0.62	0.33	0.33
	Specificity	0.47	0.87	0.75
	PosPredValue	0.97	0.03	0.03
	F1	0.76	0.05	0.06
	F2	0.67	0.10	0.12
	G-Mean	0.49	0.53	0.49

R (positive), IR (negative), PR (objective)

Table 3.4 : Confusion matrices of fine-grained filtering processes over the ATA data.

(a) STN by score		Truth		
		R	IR	PR
Prediction	R	86	6	1
	IR	31	3	4
	PR	47	6	10
	NA 44%	98	4	14
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 43%	Sensitivity	0.52	0.20	0.27
	Specificity	0.77	0.77	0.70
	PosPredValue	0.93	0.07	0.07
	F1	0.67	0.10	0.11
	F2	0.57	0.14	0.17
	G-Mean	0.51	0.37	0.41
(b) STN by polarity label		Truth		
		R	IR	PR
Prediction	R	68	4	0
	IR	29	2	9
	PR	67	9	6
	NA 44%	98	4	14
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 37%	Sensitivity	0.42	0.13	0.40
	Specificity	0.87	0.79	0.58
	PosPredValue	0.94	0.05	0.07
	F1	0.58	0.07	0.12
	F2	0.47	0.10	0.21
	G-Mean	0.53	0.30	0.44
(c) LOA by score		Truth		
		R	IR	PR
Prediction	R	162	14	22
	IR	36	1	6
	PR	27	3	0
	NA 22%	37	1	1
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 65%	Sensitivity	0.72	0.17	0.00
	Specificity	0.22	0.83	0.89
	PosPredValue	0.82	0.07	0.00
	F1	0.77	0.10	NaN
	F2	0.74	0.13	NaN
	G-Mean	0.24	0.36	0.00
(d) Naïve Bayes		Truth		
		R	IR	PR
Prediction	R	233	3	10
	IR	18	15	6
	PR	11	1	13
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 87%	Sensitivity	0.89	0.79	0.45
	Specificity	0.73	0.92	0.96
	PosPredValue	0.95	0.39	0.52
	F1	0.92	0.52	0.48
	F2	0.90	0.65	0.46
	G-Mean	0.78	0.85	0.65
(e) Neural Network		Truth		
		R	IR	PR
Prediction	R	191	3	9
	IR	30	14	8
	PR	41	2	12
Overall Statistics				
	Class	R	IR	PR
Overall Accuracy 61%	Sensitivity	0.73	0.74	0.41
	Specificity	0.75	0.87	0.85
	PosPredValue	0.94	0.27	0.22
	F1	0.82	0.39	0.29
	F2	0.76	0.55	0.35
	G-Mean	0.71	0.79	0.59

R (positive), IR (negative), PR (objective)

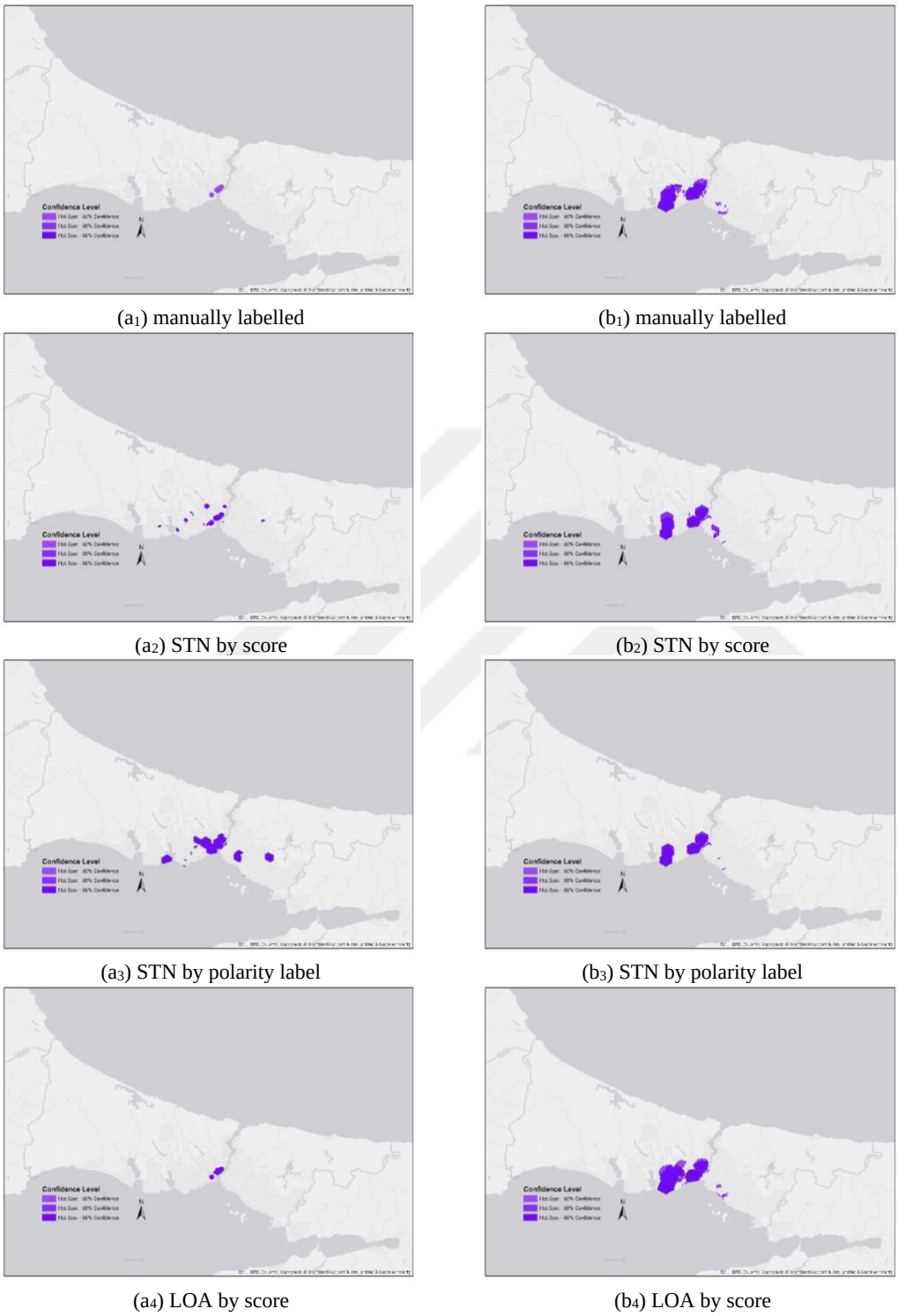
3.4.5 Spatial interpretation over fine-filtered SMD

As noted above, this study explores the fine-grained content filtering used to produce more reliable maps for specific domains. Each textually filtered dataset is overviewed with the use of confusion matrices, and this part of the study considered the textual accuracy of the tweets that can be used for a disaster domain. However, the spatial accuracy of the tweets needs to be determined in order to provide full reliability in terms of the sentiments and the location of the tweets. To determine the spatial accuracy, this study compared the manually-labelled and automatically-filtered data in the spatial context. There are diverse spatial clustering algorithms (Han et al, 2001) and methods for spatial event detection. The results vary in terms of the algorithm selection and the methodologies used (Ozdikis et al, 2017). Given the conflicts revealed using different clustering techniques and pre-specified parameters (such as the number of the clusters and required minimum number for each cluster), a Getis-Ord spatial clustering algorithm (Getis and Ord, 2010; Ord and Getis, 1995) was chosen for each dataset to compare spatial variances due to the different filtering methodologies applied previously.

While performing Optimized Hot Spot Analysis using ArcMap (Scott and Janikas, 2009), the cell size is defined as 500 metres. This represents the street-level resolution (Middleton et al, 2018) necessary to allow a fine-grained analysis. To use the connectivity capacity of a clustering lattice shape (Birch et al, 2007), hexagon polygons were selected in the aggregation method. The borderline of the city of Istanbul - where both terror events occurred - was taken as the analysis boundary. The outcomes of the analysis gave an explorative method for identifying the differences or similarities between the filtering methods and manually-labelled data (Figure 3.12 (a_1), (b_1)). Manually-labelled data hotspots for BVA (a_1) and ATA(b_1) are taken as ground truth without concerning the credibility of the data posted in social media. Explorative comparisons of hotspots could be assessed in terms of a similarity of cluster locations, cluster sizes, cluster numbers and distances between clusters.

For the ATA event, LOA (a_4) looks more similar to the truth map than all the others, while it was the second most accurate one in terms of filtering after the use of Naïve Bayes. Interestingly, when the cluster size is used, the clusters of Naïve Bayes filtering are several times larger than the ground truth. Even though, in terms of cluster location, Naïve Bayes filtering gives an accurate result, since it includes the base clusters inside,

in terms of cluster size, the level of accuracy does not meet expectations, whereas Naïve Bayes filtering comes first in terms of filtering accuracy.



(a_1) manually labelled

(b_1) manually labelled

(a_2) STN by score

(b_2) STN by score

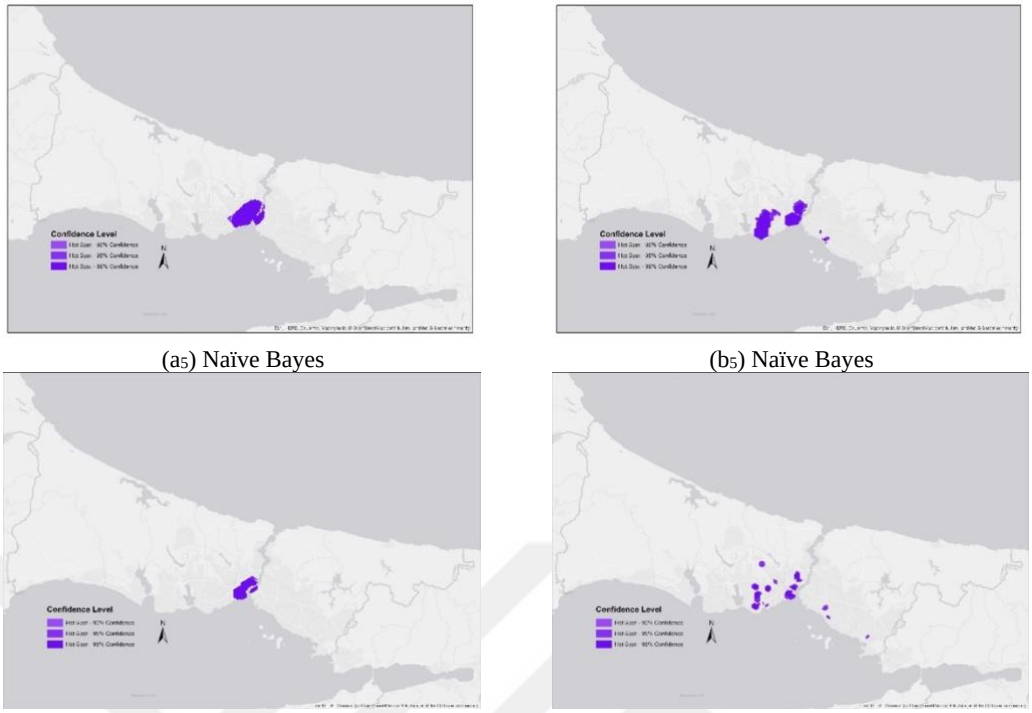
(a_3) STN by polarity label

(b_3) STN by polarity label

(a_4) LOA by score

(b_4) LOA by score

Figure 3.12 : Optimized Hot Spots over previously filtered data (a_i) BVA dataset; (b_i) ATA dataset.



(a_5) Naïve Bayes (b_5) Naïve Bayes

(a_6) Neural Network (b_6) Neural Network

Figure 3.12 (continued) : Optimized Hot Spots over previously filtered data (a_i) BVA dataset; (b_i) ATA dataset.

For the BVA event, both STN filtering hotspots were seen to be scattered to several distant locations, and STN by score (b_2), LOA (b_4) and Naïve Bayes (b_5) filtering display a rather more similar clustering pattern with manually-labelled hotspots (b_1). However, clusters generated by the data filtered with STN by polarity label (b_3) present moderate convergence to the base clusters (b_1).

The similarity rates of hotspots are listed for each event, based on their similarity indices (Table 3.5). In respect of this, the results vary in terms of the chosen indices and, at some points, the results of other indices converge to the Giz Index. There are several inferences that can be made from this. The first one is that the Jaccard Index (JI) performs well when the intersection converges to the size of the hotspots in either the ground truth or the filtered maps (Figure 3.12 (a_4), (b_6)). Secondly, since the Sorenson Index (SI) formulates the intersection as double weighted, it converges to the Giz Index when the intersection area is significantly high (Figure 3.12 ($a_{5,6}$), ($b_{3,4,5}$)). Kulczynski designates the same importance to the intersection part ratio to either ground truth and filtered map, and converges to the Giz Index, with the intersection ratio being similar in both maps (Figure 3.12 ($b_{2,4,5}$)). Therefore, it does

not converge with the Giz Index score when the intersection area is disproportionately distributed in the truth and filtered maps as shown in Figure 3.12 ($a_{2,5}$). Additionally, the Ochai Index considers the intersection ratio exponentially to total hotspot areas in both the ground truth and the filtered maps (Figure 3.12 (a_6), ($b_{2,4,5}$)). This exponential weight creates a bias in terms of the size of the intersection without bothering about the distance between the non-intersected areas. The Giz Index calculates the similarity score in terms of all the required aspects (cluster size, spatial proximity, spatial intersection area and non-intersected area) for the comparison of incidence maps.

Table 3.5 : Maps similarity rates for **(a)** BVA; **(b)** ATA.

(a) BVA	JI	SI	KI	OI	GI	(b) ATA	JI	SI	KI	OI	GI
STNScrPred	0.74	0.60	0.71	0.65	0.50	STNScrPred	0.51	0.68	0.74	0.74	0.78
STNPlrPred	0.12	0.19	0.56	0.35	0.25	STNPlrPred	0.53	0.69	0.78	0.76	0.68
LOAPred	0.90	0.64	0.96	0.96	0.85	LOAPred	0.58	0.74	0.83	0.82	0.78
NBPred	0.17	0.25	0.57	0.38	0.28	NBPred	0.55	0.71	0.81	0.80	0.76
NNPred	0.32	0.48	0.41	0.54	0.51	NNPred	0.33	0.50	0.40	0.56	0.31

The similarity threshold is identified as 0.7 for the use of filtering and mapping techniques. With regard to this, STN by score (STNScrPred) has a threshold of more than 0.70 similarity according to the JI and KI, while for the LOA (LoaPred) it has a threshold of more than 0.90 according to JI, KI, OI and GI for the BVA event (Table 3.5 (a)). For the ATA event, with the exception of NN, all Hot Spot maps displays have a reasonable similarity according to diverse indices (Table 3.5 (b)).

However, the Giz Index is accepted as the most appropriate similarity index for this study as it embodies spatial similarity, unlike the Jaccard, Sorensen-Dice, Kulczynski, and Ochai Indices. The results of the Giz Index for both events support explorative inferences, with quantitative measures of between 0 and 1, by considering both the size of spatial clusters and their proximity. This is because size and location accuracy are important for studies including spatial analyses, such as disaster management involving such an analysis. The index can provide quick and automatic spatial interpretation of the analyzed data. When it comes to the Giz Index, which allows us to evaluate the data spatially, it offers sufficient similarity for both the BVA events with LOA, and the ATA event with LOA, STN by score and Naïve Bayes techniques.

3.4.6 Outcomes

As previously mentioned, the proposed methodology of this study can also be used for different domains such as elections, natural disasters, and marketing. As an example, if the domain is an election, the following procedures can be applied to map the reactions of social media. The first step is to correctly filter the SMD. For this step the user should select the keywords for the election event such as politics, party, parliament, deputy, name of the parties, name of the candidates, election, referendum etc. Following this first step the filtered SMD will include non-relevant or partially relevant tweets due to the use of homonyms and metaphors. To filter out the non-relevant and partially relevant tweets, depending on the number of available tweets, a training dataset should be spared and labeled manually as being relevant, non-relevant and partially relevant. This step can be automatically done by using web platforms such as Kaggle and Mechanical Turk (Cieliebak et al, 2017; Sorokin and Forsyth, 2008). This manually or web-based labeled dataset is used to train a Naïve Bayes Classifier, and this classifier is used to fine-grain filter the SMD. After the automatic fine-grain filtering, the relevant resulting labeled data can be used to create hotspot maps of social media reaction, changes in reaction maps, and the thematic and spatial distribution of reactions by using the Getis-Ord* algorithm. This map can show most and least favorite candidates, parties, promises, and the complaints of the electors in different spatial regions.

Temporal reaction maps can be compared by using the Giz Index to evaluate differences in terms of the election campaign based on the candidates or parties. The reaction maps of the candidates or parties before and after meeting times can also be compared by using the Giz Index to investigate the reaction of the electors to the promises made by the candidates.

Another aspect of this study's methodology relates to event-based mapping for the determination of the size and distribution of the event, to find the most dangerous, risky or safest routes for evacuation that were used by the tweet owners or the most secure locations. An example of such a case is the use of this methodology during and following an earthquake event to create an incidence map. The first step is again to pre-determine the keywords before the event. Then, during and following the earthquake event, the SMD is pre-filtered. To fine-grain filter the SMD following the pre-filtering, a training dataset should be spared for training the classifier, either by

manually labeling relevant, non-relevant, and partially relevant data, or by using web platforms. Following the labeling, the training dataset is used to train Naïve Bayes Classifiers. After the supervised Naïve Bayes classification, unsupervised topic modelling techniques such as latent semantic analysis (LSA), probabilistic LSA (pLSA) or latent Dirichlet allocation (LDA) can be used to categorize the incident reporting tweets (Hu et al, 2012; Sridhar and Vivek, 2015). If the SMD is open for manipulation based on bot accounts multiple tweets, the data can be refined by the location and the username, and the manipulation sources and affect can consequently be minimized. Following this step, a Getis-Ord* algorithm can be used to create incidence maps of most of the collapsed buildings that exist, most of the people who are waiting for a response, or the most secure locations or routes for gathering or evacuation. The incidence maps can be compared with the emergency calls, pre-earthquake risk maps and response plans by using the Giz Index. In this way, previously estimated damage and risks can be validated, and the current situation affecting the information flow can be controlled by using the proposed methodology. This difference can be used to optimize the response plans rapidly following an earthquake.

3.5 Conclusion

The social media data generated by billions of bionic sensors throughout the world, and by nearly half of the total population of Turkey, is crucially significant as a data source, during and after a disaster. This study determines the value of fine-grained filtering techniques in the terror domain with regard to the Turkish language. The point of this study, in terms of using fine-grained filtering, is to handle directly-relevant data in relation to a specific domain, which in this case was selected as terror.

The first outcome in the context of this study is the processing of Turkish tweets related to two terror attacks with current subjectivity lexicons. The current Turkish subjective lexicons are utilized for filtering. Since these libraries offer only a low success rate, two machine learning classifiers are integrated into the analysis instead. Consequently, it is identified that those lexicons are partially adequate when it comes to filtering out non-relevant positive or objective content. Instead of these, a Naïve Bayes classifier is trained using other similar event data, and this classifies test datasets with over 80% accuracy. As a result of this study, the highest success is achieved with the use of Naïve

Bayes techniques, while the LOA achieves the second-best success rate. Although Naïve Bayes provides the best results in terms of classification, it requires trained datasets for each domain. Such trained data might not be easily collected as it depends on major disaster events. For example, Istanbul, as one of the most crowded cities in the world, is expecting a major earthquake. However, until a damaging earthquake occurs it will not be possible to gather training data, and this reality challenges training the classifier for all the sub-domains of disaster. Even general reactions might be common for all types of the disaster domains such as 'God's mercy' or 'wishing a speedy recovery to the wounded'. If the study case includes only a specific domain such as terror within the range of disaster domains, there needs to be an event to gather and train the data. Further studies are planned to involve other disaster domains such as fire, flood and earthquake, in order to build classifiers based on the Turkish language.

The spatial accuracy of each tweet is another aspect that needs to be considered to ensure fine-grained incidence mapping with a high degree of accuracy, though the study covers attribute filtering for mapping purposes. The second outcome of the study is the review of indexing methods to compare the spatial distribution of filtered data (prediction) versus manually-labelled data (truth). This study achieved 85% accuracy with regard to text and spatial data mining with the use of the chosen techniques relating to social media data. Such finely-filtered spatially credible data could serve as auxiliary data to allow stakeholders to rapidly determine the situation following a disaster event, as notification of an emergency situation, as a request for help or for a public gathering, or notification of hazardous locations.

A case study with two events compared with 3 lexicon-based and 3 machine learning based sentiment analyses to assess the methodology to quantify impacts of filtering techniques on spatial reliability of the social media data. First two is the based on STN (Dehkharghani et al, 2016) library resulting as; score based and label-based analyses. The third one is the lexicons of (Ozturk and Ayvaz, 2018) which was mentioned as LOA within the study. There were also 3 machine learning based sentiment analyses ran within the study as; Naïve Bayes Classifier and Neural Network Classifier with 1 hidden layer, 2 hidden layers and 3 hidden layers separately. The last machine learning based analysis is the Support Vector Machine with polynomial kernels, radial kernels, and sigmoid kernels separately.

Following the assessment of the filtering methodologies, the SMD analysis results were evaluated by using the similarity indexes and the ground truth data. Well-known similarity indexes were compared with the Giz index, which was developed within the study. The comparison results of the similarity indices reviewed in this study show that the indexing methods score non-/intersection as binary (0-1) without considering spatial relationships. When spatial relationships are not considered within the SM analyses, it also increases the debate on the reliability of the SMD. The biggest obstacle with regard to using social media data in scientific analyses is the reliability of the SMD. There is an insufficient number of reliability studies on the textual and spatial context of SMD. This is where this study provides an index for spatial reliability and evaluates the textual contents of SMD by using text-based filtering over tweets for a specific domain. In this way, a numerical evaluation of the SMD can be generated by use of the proposed methodologies.

The results of this research provide a new spatial similarity index in the form of the Giz Index, and offers a combination of existing and newly-developed techniques and methods to filter Twitter data for a domain with a high degree of relevance, and to produce an event map for that domain, then determine the spatial accuracy of the map with ground truth data.

It is also important to note that there are few studies in this area with regard to non-English languages. This study offers a methodology that allows us to work with agglutinative languages, especially Turkish, where the majority of the population is not tweeting in English.

As a result, this study provides a spatial similarity index with regard to the community, which deals with spatial intersection, proximity and size together. It is the first study to spatially analyze filtering techniques with regard to SMD, and offers a method that does not only stick to domain consistency and semantic relevance, but also takes into account the spatial reliability of the SMD in conjunction with them.

4. CITIZENS' SPATIAL FOOTPRINT ON TWITTER: ANOMALY, TREND AND BIAS INVESTIGATION IN ISTANBUL[3]

4.1 Abstract

Social media (SM) can be an invaluable resource in terms of understanding and managing the effects of catastrophic disasters. In order to use SM platforms for public participatory (PP) mapping of emergency management activities, a bias investigation should be undertaken with regard to the data related to the study area (urban, regional or national, etc.) to determine the spatial data dynamics. Thus, such determinations can be made on how SM can be used and interpreted in terms of PP. In this study, the city of Istanbul was chosen for social media data research area, as it is one of the most crowded cities in the world and expecting a major earthquake. The methodology for the data investigation is: 1- Obtain data and engage sampling, 2- Identify the representation and temporal biases in the data, and normalize it in response to representation bias, 3- Identify general anomalies and spatial anomalies, 4- Manipulate the trend of the dataset with the discretization of anomalies, and 5- Examine the spatiotemporal bias. Using this bias investigation methodology, citizen footprint dynamics in the city were determined and reference maps (most likely regional anomaly maps, representation maps, time-space bias maps, etc.) were produced. The outcomes of the study can be summarized in four steps. First, highly active users generate the majority of the data and removing this data as a general approach within a pseudo-cleaning process means concealing a large amount of data. Second, data normalization in terms of activity levels, change the anomaly outcome resulting from diverse representation levels of users. Third, spatiotemporally normalized data present strong spatial anomaly tendency in some parts of the central area. Fourth, trend data is dense in the central area and the spatiotemporal bias assessments show the data density varies in terms of the time of day, day of week and season of the year. The

[3]This chapter is based on the article: Gulnerman, A.G., Karaman, H., Pekaslan, D., Bilgi, S., 2020: Citizens' Spatial Footprint on Twitter: Anomaly, Trend and Bias Investigation in Istanbul, ISPRS International Journal of Geo-Information, 9, 222. doi: https://doi.org/10.3390/ijgi9040222

methodology proposed in this study can be used to extract the unbiased daily routines of the social media data of the regions for the normal days, and this can be referred for the emergency or unexpected event cases to detect the change or impacts.

Keywords: volunteered geographic information; social media; public participation; spatiotemporal bias

4.2 Introduction

Over the past two decades, public participatory (PP) mapping has evolved rapidly from paper to digital mapping (Ball, 2002; Hall et al, 2010; Sieber, 2006). Social media (SM) platforms that provide huge volume of crowdsourced data to digital mapping are not regarded as an appropriate participatory mapping resource, even though SM can be used as a pioneering platform that increases individual participation rates for PP mapping with its features of large data production capacity and uninterrupted data collection platforms. Global Navigation Satellite Systems' (GNSS) antennas in smart devices, and public use of these devices, enable and foster location-based crowdsourced applications. The data is generated by the users of these applications and is referred to as Volunteered Geographic Information (VGI) (Elwood et al, 2012; Goodchild, 2007a). The users can be thought of as unconscious volunteers for social media (SM) VGI, as deliberate volunteers for peer production VGI, and as public participators for in citizen science based VGI (Gulnerman et al, 2016; Hecht & Shekhar, 2014; Hecht and Stephens, 2014). The way of producing these forms of VGI is referred to as neo-geography in that it adopts neo-geographers (i.e. volunteers) who contributes to mapping activity without being expert (Goodchild, 2009). This inexperience with regards to data production is questioned in the context of data quality (Ballatore and Jokar Arsanjani, 2018; Haklay, 2010; Mooney et al, 2010), demographic bias (such as, gender, socioeconomic and educational aspects) (Gardner and Mooney, 2018; Stephens, 2013) sampling bias (referring to volunteer sampling) and its impact on the generated data (Brown, 2017; Haklay et al, 2010).

In their very first form, citizen science projects in the very first forms were carried out with the use of paper maps (Ball, 2002). However, with the technological developments in computer and web sciences, nowadays they are mostly carried out with the help of a range of online platforms (Scistarters, 2017; Ushahidi, 2017; Zooniverse, 2017) designed for collecting data for citizen science purposes. These

platforms are designed to collect data within a limited time period for specified purposes. On the other hand, there are also dedicated webpages deployed for local citizen science projects such as DYFI (Did You Feel It) (Wald et al, 2012). DYFI served by the USGS, collects data from volunteers with regards to how intense they feel an earthquake, in order to show the extent of damage and shaking intensities on a map. Although the project is designed and structured to collect and process data, the count in terms of volunteers' response responses to individual earthquakes (the participation rate) to each earthquake is pretty low (USGS, 2019). Although the project has been replicated for different countries such as New Zealand, Italy and Turkey (Tarhan et al, 2013; Wald et al, 2012) there is still not any organized approach by the relevant authorities (Kocaman et al, 2018).

SM platforms, although not seen as the proper way to organize citizen-based projects, still have a high and continuous data serving capacity around the world involving 3 billion of users (Statista, 2020). In fact, SM with its wide, continuous, active data collection and serving capacity can be used as a pioneering platform for carrying out citizen-based projects especially for monitoring out of the ordinary events such as multi-emergency circumstances in big cities. However, bias in SM data can be seen an obstacle with regard to such projects. In order to use SM data as a citizen-based monitoring system, the data georeferenced in a particular area (such as a city, region or country) should be pre-assessed. In this way, ways in which to use and interpret the SM data as a tool with regard to city monitoring can be inferred.

4.2.1 SMD studies on emergency mapping

SMD has already in use for disaster management for more than a decade. (Sakaki et al, 2010) presents a comprehensive literature on the functionality of SM in terms of the disaster management phases. The very first example of event detection with SMD was conducted by (Sakaki et al, 2010). Social media was also considered for disaster relief efforts by (Gao et al, 2011) and (Muralidharan et al, 2011), for crisis communication by Acar and Muraki (2011) and McClendon and Robinson (2012), and for evacuation ontology by Ishino et al. (2012) and Iwanaga et al. (2011).

Most of the former and latter studies have focused text-based filtering at first for detecting an event (Bruns and Liang, 2012; Mendoza et al, 2019; Wang et al, 2016; Zou et al, 2019). The filtering techniques were mostly used for a limited number of

keywords related to a disaster domain (such as; hurricane, flood, and storm for meteorological disasters) (Zou et al, 2019). In respect to that, this kind of studies has selection bias due to determined keywords (Leetaru et al, 2013). This might not be a problem for coarse-grained spatial analyses due to an abundance of data however; this may lead cause detection problems for the local events. Yet the studies are mostly basing on an event type instead of being a comprehensive monitoring system to detect any disastrous event anomalies. In addition to that, the spatial grain of the detection analyses are mostly coarse as county or city level (Mendoza et al, 2019; Zou et al, 2019), even the studies are focusing the spatial consideration at first (Leetaru et al, 2013).

Historical data exploration plays an important role to comprehensively monitor unusual events in a fine-grain spatial level within a city (Leetaru et al, 2013; Middleton et al, 2014). That is why SMD should be assessed in terms of anomalies, trends, and bias. In this way, citizen footprints on social media can be interpreted in several ways as base maps for further local event detection investigations. The most operational use of the proposed method can be on emergency mapping because of rapid succession and the ability to compare the difference with the daily life trends. Since, phases of emergency management require rapid real-time data on the region of interest to compare the ongoing situation with the preparedness plans.

4.2.2 Aim and region of the study

In this study, the aim is to propose a methodology for bias investigation in order to reveal citizens' footprints in a city, and to produce reference maps (most likely regional anomaly maps, representation maps, spatiotemporal bias maps, etc.). The city of Istanbul was chosen as the case study area as it is one of the biggest cities in the world with 18 million inhabitants, and one which expects a major earthquake that could possibly have a catastrophic impact on the city (Karaman et al, 2008; Karaman and Erden, 2014). Twitter platform is used as the data source, since it is one of the most commonly-used social media network for spreading the information all over the world (Sloan and Morgan, 2015; Statista, 2019). Such data is referred to as Social Media Data (SMD) in this paper. With regard to the investigation of the SMD, the methodology includes the following steps: data acquisition and data tidying, determination of the representation and temporal bias in the data, data normalization for removing user

representation bias, detection of anomalies in non-spatial and spatial data, the discretization of anomalies and the production of a trend map, and the investigation of spatiotemporal bias. The data investigation outcomes in this study are discussed from the perspective of citizen-based event mapping with the use of SM data. In this way, the assessment techniques with regard to SM data is presented in this study for the benefit of the citizen-based geospatial mapping capacity building.

4.2.3 Bias in SM-VGI

Nearly half of the world population are unrepresented in SM due to internet censorship or access unavailability (Statista, 2020). Consequently, this study of bias in social media data (SMD) starts with a consideration of the inadequacy of the technological and political infrastructure. Moreover, the usage rate of smart devices and computers affects the representation rate of societies and individuals. In addition to the representation of societies may not be equal which is mostly explained in terms of demographic (age, education, social status) differences. For several reasons, some parts of a society might be over-represented while other parts may be under or not represented at all (Basiri et al, 2019; Haklay, 2010). However, determining demographic bias is mostly not possible due to the lack of availability of volunteers' personal data in VGI (Basiri et al, 2018; Brown, 2017). Additionally, volunteers of the platform might not even be a person, in that they can be a bot, a staff team member (who embodies and promotes a company) and/or a troll (a fake account).

Basiri et al. (2018) suggest that there are more than 300 types of bias, and that crowdsourced data might tend to have some of those. Since volunteers are contributing directly without being asked to participate, SMD do not include *"selection bias"*. However, the volunteers diversely show *"representation bias"* due to their immense activity rate. Also, population density possibly creates *"systematic bias"* over space. Due to the reputation of places, volunteers tend to share popular locations more, to flaunt and to be visible at the same page with others. This is referred to as *"Bandwagon bias"* and as which affects the spatial distribution of VGI. "*Status-quo bias"* is also a reflection of the demographic background of volunteers over space, and is seen as specific types sharing a point of interest (Basiri et al, 2018). While Bandwagon and Status-quo biases plead to spatial bias, the temporal dimension creates varying patterns with regard to changing activities or participation rate s corresponding to the time of

day, day of the week and season of the year. This changing trend is referred to as "*spatiotemporal bias"* and entails misinterpretation in the case of a comparison of improper temporal slices.

There have been number of attempts to identify spatial patterns, trends and biases with regard to the SMD. Li et al. (2013) conducted a research on Twitter and Flickr data at the county level in order to understand users' behavior as a result of demographic characteristics. The study also offers some exploratory graphs about the number of tweets over time, and presents tweet density maps that are normalized by the population density at the county level. Another study is about understanding the demographic characteristics of users who enable location services on Twitter (Sloan and Morgan, 2015). The study offers strong evidence based on demographic effects on the tendency of enabling geo- services and geotagging. Lansley and Longley, (2016) searched for the dynamics of the city of London using topic modelling and quantified the correspondence of topics with the users' characteristics and location. Arthur and Williams (2019) conducted a research to identify regional identity and inter-communication between cities. The researchers found that regional identity that is quantified by text similarity and sentiment analysis of posted tweets in terms of several UK cities. Malik et al. (2015) conducted research to find the relationship between the census population and the number of geotagged tweets using statistical tests. They found that there were no impacts of population on tweets density. However, they did identify several other impacts such as the income level of the population, being in a city center, and the age of the population. Another study with regard to the user bias in terms of the tweeting frequency of users, proposed the removal of the top 5% of active users from the data to avoid such biases (Tsou et al, 2017).

In this respect, the studies carried out mostly focused on the demographic background of the users, the relationship between population and tweets density, representation bias, or topic variances over coarse-grained space. However, those were not searching for a year based fine spatial data pattern, and for biases that can allow the monitoring a city with a better interpretation of spatiotemporal data. However, this study is designed to present spatiotemporal variances with regard to representation diversity, and anomalies and trends in the data, without blocking or removing any users' data that is a commonly adopted way of previous studies for data cleansing.

4.3 Materials and Methods

The methodology of this study is presented in five subsections and the conceptual flow of methodology can be followed from Figure 4.1. In the first subsection (4.3.1), details of data acquisition techniques and data tidying steps are explained. In the second subsection (4.3.2), the data investigation methodology in terms of users' activity levels and temporal levels are introduced. In addition, the application of user-weighted normalization techniques is introduced to investigate the impact of users' activity levels on the temporal data variation. In the third subsection (4.3.3), details of the investigation of data anomalies are presented in two stages anomaly detection over non-spatial data, and anomaly detection over spatially-indexed data. In the fourth subsection (4.3.4), the methodology involved in obtaining regular data is explained in order to produce spatially-indexed overall trend data and a map. In the fifth subsection (4.3.5), bias assessment details are incorporated into the methodology flow. Bias investigation in terms of temporal levels is explained step-by-step in this last part.

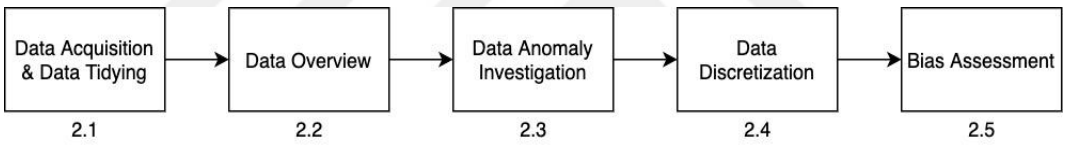


Figure 4.1 : Conceptual Data Investigation Flow in the Subsections.

4.3.1 Data acquisition and data tidying

The data flow of this part is composed of the following steps - data downloading, storing, sampling and tidying (Figure 4.2). Geo Tweets Downloader (GTD) (Gengec, 2016) is chosen as the downloading software, since this study aims to monitor tweeting pattern within a spatial bounding box. GTD is a software that uses Twitter APIs to download georeferenced tweets and ingests this data into PostgreSQL in real time. GTD has acquired data during the year 2018 within the bounding box of Istanbul City. There were several interruptions such as electricity and internet cuts in the downloading server during this data acquisition process that lasted for one full year. Therefore, the acquired data has been sampled into weeks. Data continuity from Monday to Sunday inclusive was determined as the only sampling rule, and starting from the first week of each month, each hour was checked as to whether or not there was a missing data. Based on this, the data was composed of 12 complete selected weeks belonging to each month of the 2018 year.

Data was tidied with the addition of three temporal level columns for further investigation in the following sections. The first level (timeLevel1) shows four different time intervals of the day; night (00:00-06:00), before midday (6:00-12:00), after midday (12:00-18:00), and evening (18:00-00:00). The second level (timeLevel2) shows the day of the week from Monday to Sunday inclusive. The third level (timeLevel3) shows the month of the year from January to December inclusive. In the database time level values are represented by an integer value. On this basis, timeLevels 1,2 and 3 have 4, 7 and 12 integer values respectively. In addition to these time level columns, a time vector column was calculated with the formula given in Figure 4.2. According to this calculation, the timeVector has values from 1 (night, Monday, January) to 336 (evening, Sunday, December).

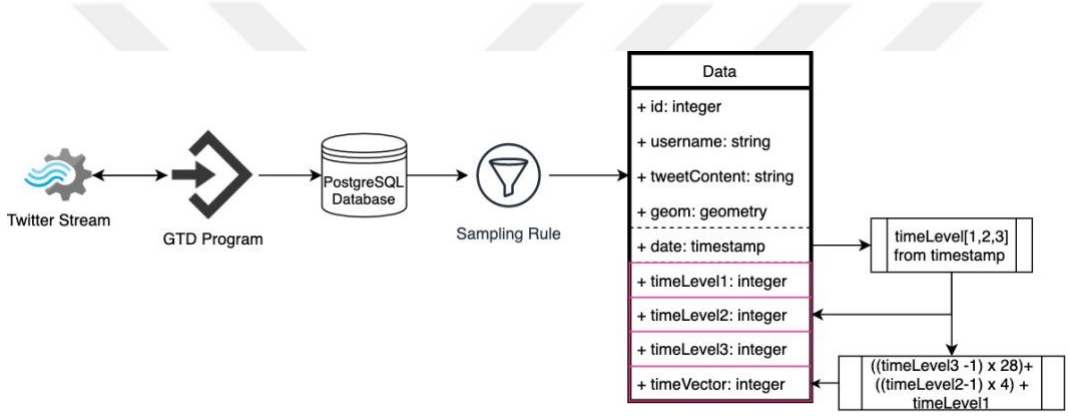


Figure 4.2 : Data downloading and tidying flow.

4.3.2 Data overview

Data investigation was composed of a user representation level search, the tweet count variation in terms of time level, and a normalized tweet count in terms of time level. The investigation started with data generators representation levels that may cause noisy weights over data. The activity level of each user was determined by using one of the most common k-means clustering technique on the overall users' activities. In terms of the activity level decision, 1- data was grouped by username, 2- a tweet count for each user was calculated, 3- the min, mean, standard deviation values of tweet counts were calculated (Figure 4.3). The histogram of the tweet count was not well represented since it was highly right skewed. However, a summary of tweet count is as follows:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.
1.00	1.00	4.00	53.91	17.00	4378.00	185.474

4- As the tweet number 1 is associated with many users in the overall dataset, it has been separated as the first cluster. Since the data is not normally distributed, the k-means clustering is applied on the remained dataset by separating it into 3 clusters. In the k-means clustering implementation, we followed the traditional approach as listed below.

1. 4 is chosen to be clusters number
2. Place the centroids c_1, c_2, c_3 and c_4 randomly
3. Repeat steps 4 and 5 until convergence or until the end of a fixed number of iterations
4. For each user's tweet number - find the nearest centroid (c_1, c_2, c_3 and c_4) - assign the the user to that cluster
5. For each cluster j = 1..4 - new centroid = mean of all points assigned to that cluster
6. End

   Each cluster representation can be seen in Table 4.1.

According to this, the min value 1 was specified as the single representation class of the activity level, the max value of cluster 2 (261) is bounding the upper part of the second activity class, the max value of cluster 3 (944) is bounding the upper part of the third activity class and lastly, the max the max value of cluster 4 (4378) is bounding the upper part of the fourth activity class. With the addition of the user activity levels to datasets as shown in Figure 4.3, the number of users and tweets in terms of activity level were investigated in terms of the plots in the "Results" section.

Table 4.1 : User representation level clusters detail.

Cluster	Min	Max	Average	Std. Dev.
c_1	1	1	1	0
c_2	2	261	26.68	41.18
c_3	262	944	497.95	185.44
c_4	947	4378	1393.81	397

The number of tweets in terms of time levels was also investigated in terms of the circular bar plots in the "Results" section, to make general inferences with regard to any temporal bias. The variation in the number of tweets was firstly plotted over the raw data count without any weighting. In addition, the number of tweets is normalized with the technique described below. The normalized number of tweets in terms of time level is investigated in order to discuss the impacts of the representation level to temporal data variations.

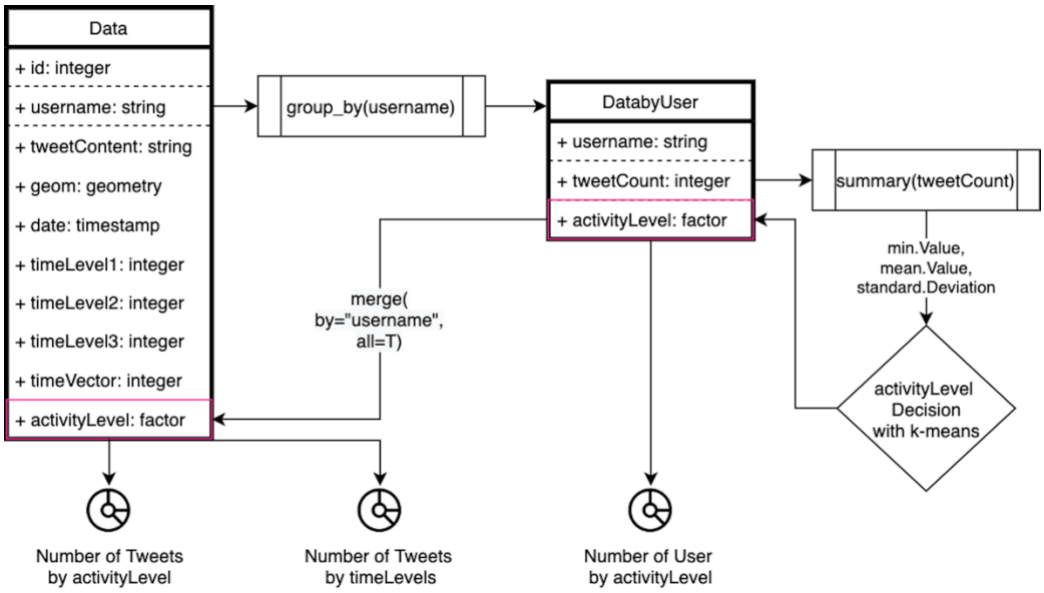


Figure 4.3 : Data investigation methodology flow.

Due to the nature of the SMD data, there is a noticeable variance in the number of tweets and user behaviors which lead to some data being "outlier", a term which is used to describe any unusual behaviors. However, it is often not possible to determine whether those outlier data are invalid or whether the data represent valid information for the overall dataset or task to achieve. Therefore, in this paper, those outlie tweet numbers/users' behaviors, which deviate noticeably from the overall data, are purposely not removed, but rather all the users are assigned weights based on their activity levels. Through the use of these assigned weights, each user is represented at various levels in the used data latter. Thus, we followed two-step normalization procedures (entitled inter-normalization and intra-normalization) to capture and handle each particular user's behavior and the tweet information that exists.

In terms of intra-normalization, each users' weight is determined based on their tweeting numbers in comparison with the overall tweet dataset. In this paper, we follow the analogy that if user A tends to send large numbers of tweets (e.g. ~100) on a normal day, this number of tweets will increase--accordingly in the event of a disaster occurrence. On the other hand, user B, who tends to send a smaller number of tweets (e.g. 1 or 0) on a normal day, will increase the number of tweets in parallel with this normal daily tweet number. Therefore, user A is assigned with a smaller weight than is user B for each tweet. This approach can allow us to cope with any discrepancy in tweeting behavior between the users. The weight determination of each user is carried

out on their overall tweet numbers in the overall dataset, and each user's tweet number is divided by the maximum (c_{max}) number of tweets.

The user weighting is implemented as follows:

$$w_i = \left(1 - \frac{u_i}{c_{max}}\right) \tag{4.1}$$

where w_i is the determined weight of the i'th user, u_i is the total tweet numbers, and c_{max} is the maximum number of tweets which were sent respectively.

After each user's weight is determined, the overall tweet number (n_i) of each user in the general pool is calculated by simply multiplying the user weight with the users' overall tweet number. Consequently, the inter-normalization is completed by allowing each user tweet's number to contribute in a 'compromising manner' to the general pool. In addition, by following this procedure, each user's contribution is utilized without being ignored or removed from the dataset. After gathering the weighted tweet numbers (n_i) from each user, in the intra-normalization, time-based tweet numbers are summed and the commonly-used cube root transformation is implemented for each time-based tweet number (4.2), and the min-max normalization is applied on this gathered ($\boldsymbol{c^t}$) tweet numbers dataset

$$c^t = \sqrt[3]{\sum_{i=1}^{N} n_i^t} \tag{4.2}$$

where c^t is the transformed tweet number on a time *t*. *N* is the weighted user numbers which are tweeted on the time *t*, and n_i is the weighted tweet numbers from the intra-normalisation. Consequently, that applying intra and inter-normalization procedures enable the representation of each user and each tweet in the final normalized dataset.

4.3.3 Anomaly in data

The anomaly investigation was carried out in two stages that were applied to both non-spatial data and spatial data. The AnomalyDetection R package (Twitter, 2015; Vallis et al, 2014) was adopted for the applications. The package was created by Twitter for anomaly detection and for visualization where the input Twitter data is highly seasonal and also contains a trend. The package utilizes the Seasonal Hybrid Extreme Studentized Deviate test (S-H-ESD) which uses time series decomposition and robust statistical metrics along with the ordinary Extreme Studentized Deviate test (ESD).

The S-H-ESD provides sensitive anomaly output specializing in Twitter data, with the ability to detect global anomalies as well as anomalies have a small magnitude, and which are only visible locally. To compute the S-H-ESD test, Anomaly Detection package provides support for the time series method and for the vector of numerical values method where the time series method gets the timestamp values as inputs while the vector method requires an additional input variable "period" for serialization. Both methods require a maximum anomaly percentage, "max_anoms" (upper bound of ESD) and the "direction" of the anomaly (negative, positive or both) (Twitter, 2015; Vallis et al, 2014).

In this study, the vector method of the AnomalyDetection package is used. The period variable is set as 28, since 7 days of data is used, and since each day is divided into 4 periods according to the hour of the given tweets. (Hochenbaum et al, 2017) experimented with 0.05 and 0.001 as the maximum anomaly percentages in their experiments. They achieved better precision, recall and F-measure values with 0.001, though with very few differences between each other. Considering the experiment setup of (Hochenbaum et al, 2017), the maximum anomaly percentage is chosen as 0.02 as the optimum value.

For the first anomaly application stage as presented in Figure 4.4, 1- data is grouped by timeVector and summarized as tweetCount, userCount and normalizedTweetCount, 2- the counts are ordered by timeVector, 3- anomaly detection is applied to the vectorized counts.

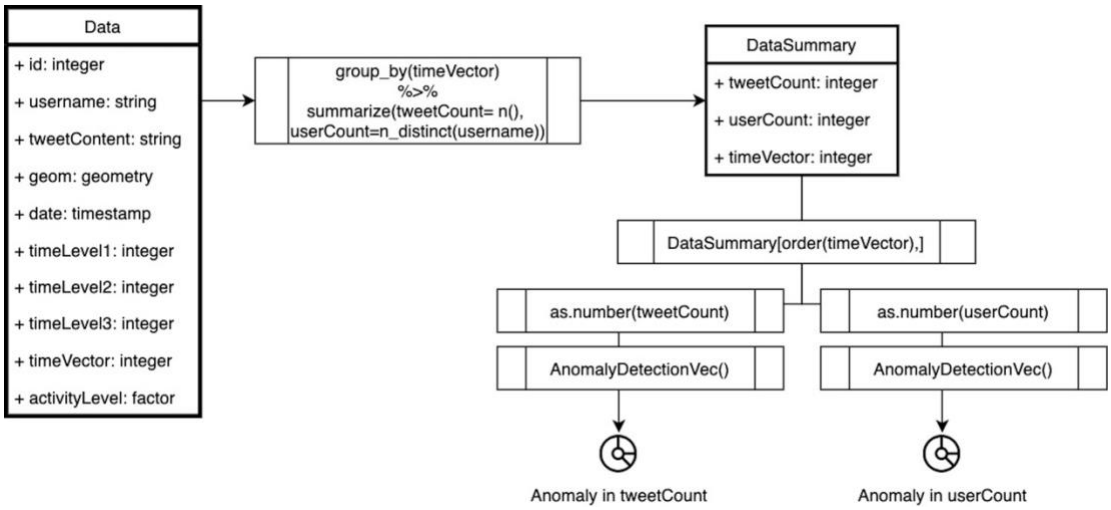


Figure 4.4 : Anomaly Detection Stage 1.

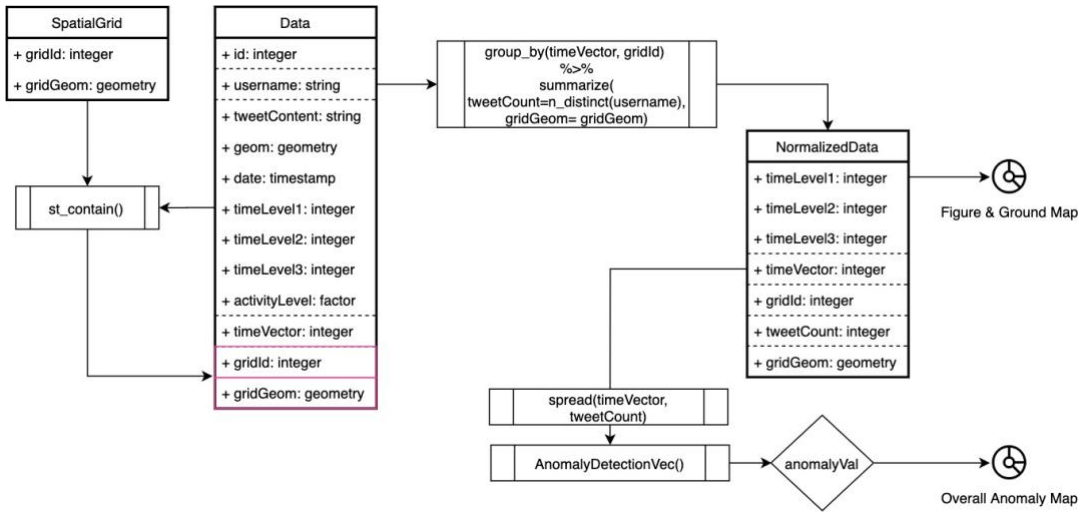


Figure 4.5 : Anomaly Detection Stage 2.

For the second stage, spatiotemporal anomaly is assessed as presented in Figure 4.5. The steps are 1- spatial grids (1 x 1 km) within the Istanbul bounding box which is spatially joined with the data, 2- data is grouped by timeVector and gridId, and summarized as tweetCount (as the distinct count of the usernames) and grid geometry, 3- a figure and ground map is visualized to display unrepresented spatial grids, 4- anomaly detection is applied to the spatially normalized tweetCount for each grid, 5- anomaly assessment is done with the normalized anomaly rate in terms of time intervals, 6- an anomaly map is visualized and spatial pattern is tested with the Moran *I*.

In the second step, tweets from the same user within the same time interval and a 1 x 1 km grid are counted as 1 tweet. This is done to avoid over-representation of a user. In the fifth step, the normalized anomaly rate is formulated (4.3) for the anomaly assessment. The anomaly and expected values that are provided by the AnomalyDetectionVec part is used with the timeVector indexes (i) in this formula and the normalized anomaly rate is calculated for each grid and timeVector pair. The overall anomaly map is produced with the sum of the normalized anomaly rate for each grid. By this normalized anomaly rate values, the most anomalous spatial grids were plotted and tested with the Moran's *I* algorithm.

$$\text{normalized anomaly rate} = \frac{anomalyValue \; - \; expectedValue}{\sum_{i(timeVector)} \text{anomalyValue} - \text{expectedValue}} \tag{4.3}$$

Moran's *I* is the measure of global and local spatial autocorrelation. Global and local Moran's *I* are utilized for this part and for the following parts of the study, in order to determine the observed anomalies, trends and temporal differences, either clustered,

dispersed or random in space. The "spdep" R package (Bivand et al, 2015) was used to calculate global and local Moran's *I* which are formulated (4.4),(4.5) based on the feature's location and the values of the features (Anselin, 1995).

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4.4}$$

$$I_i = \frac{(x_i - \bar{x})}{\sum_{k=1}^{n}(x_k - \bar{x})^2 / (n-1)} \sum_{j=1}^{n} w_{ij}\,(x_j - \bar{x}) \tag{4.5}$$

The variables are n=number of features indexed by i and j, x=spatial feature values, x=mean of x, w_{ij}= matrix of feature value weights. The values of Moran's *I* range from -1 (negative spatial autocorrelation) to 1 (positive spatial autocorrelation) and returns 0 value for a random distribution. The Spdep package provides moran.test() and localmoran() functions. In this study, both for global and local spatial autocorrelation calculations, moran.test() and localmoran() functions were used with the following arguments: x (the numeric vector of the feature attributes), listw (spatial weights for neighbouring lists that is calculated by the nb2listw function in the spdep package), zero.policy (specified as TRUE to assign zero value for features with no neighbours). The functions' return values include Moran statistics (I, I_i) and the p-value of the statistics (p.value, Pr()). A value of less than 0.05 for the p-value means that the hypothesis is accepted, and is spatially correlated for Moran's *I* (Anselin, 1995; Bivand et al, 2015). It is also taken into consideration for interpreting the results of all Moran's *I* tests in this study.

4.3.4 Data discretization and trends

In this part of the study, the trend dataset is manipulated as presented in Figure 4.6. Steps are as follows: 1- detected anomalies in the data were discretized, 2- the discretized anomalies were replaced with expected values in terms of gridId and timeVector and assigned as regular data, 3- data were grouped by gridId with the summary of the tweetCount mean value for the overall trend data, 4- an overall trend map was produced and tested with Moran I for the trend values' spatial pattern determination. The trend map so produced represented the general dynamics of the city and was also used as the reference in order to quantify the spatiotemporal bias in the next section.

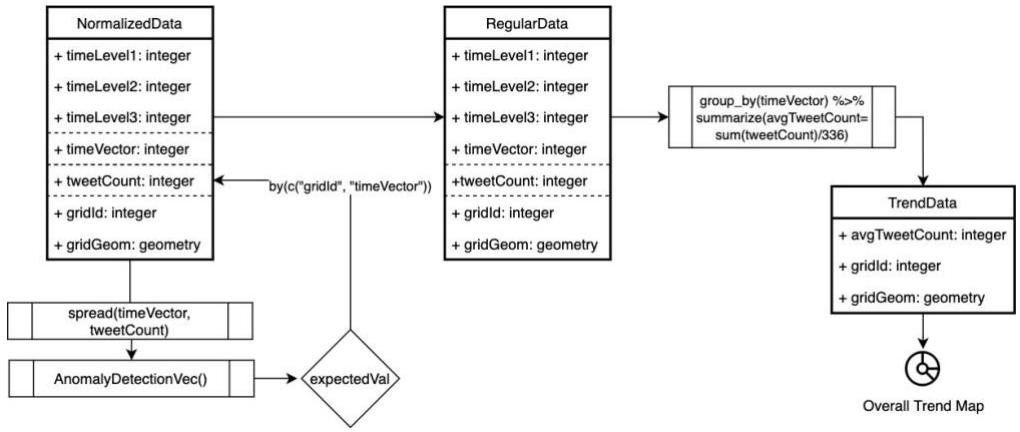


Figure 4.6 : Data replacement with expected value.

4.3.5 Spatiotemporal bias assessment

The spatiotemporal bias was assessed in terms of hour, day and seasonal levels and the flow of the assessment was displayed in the Figure 4.7. The assessment steps are as follows: 1- data was divided into sub-datasets in terms of time levels as 4 sub-datasets (night, bmidday, amidday, evening) for timeLevel1, 7 sub datasets (from Monday to Sunday) for timeLevel2, 4 datasets (winter, spring, summer, autumn) for timeLevel3, 2- these sub-datasets were grouped by gridId and summarized as average tweetCount, 3- each grouped and summarized dataset was assessed with a comparison between the avgTweetCount of each grid in the sub-data and the trend data, 4- Maps of the bias assessment were visualized and tested with the Moran I for spatial pattern investigation.

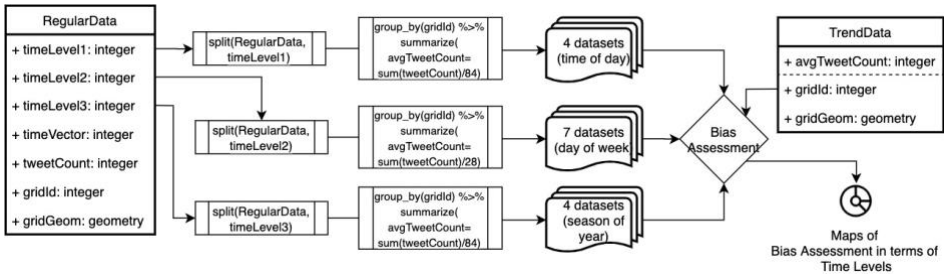


Figure 4.7 : Bias assessment flow.

In the bias assessment part of the flow, the comparison was handled by taking the average tweet counts' difference in sub-datasets and trend data. This differences in value were plotted in five classes; two classes (low, less) for negative values, a trend class for 0 values, two classes (more, high) for positive values. These values were tested with the Moran I for quantifying the spatial correlation of the values.

4.4 Results

Data acquired and sampled for this study covers over 4 million tweets generated by nearly 76 thousand of volunteers. The most active volunteer has 4378 tweets, while one third of all volunteers have just one tweet within all data. Average tweet count per user is 54 with 186 standard deviation. This reveals that activity of volunteers are disunited, and the most active group is highly over represented. This condition requires scrutinization to understand whether those groups are more active due to the unordinary situations or as their general behavior. As initial step to explore data, user's activity levels are classified by depending on the k-means clustering method instead using min, mean, and standard deviation values of user's tweet numbers as explained in the methodology. In respect to this, users' representation levels (R.L.) are classified as a single (1), second level representation (2), third level (3) and fourth level (4) as the highest active class. The percentage of users per representation levels (a) and the percentage of tweet amounts corresponding to users' representation levels (b) are illustrated in Figure 4.8 as chord diagrams. The diagrams present, nearly 90% of users represent themselves one time or less than 262 times, just the opposite their total representation in data equals to less than 30%. This initial analysis uncovers the reality in diverse representation levels of users and this variation points to representation bias. Many studies in literature omit the data come from over-represented groups in order to decrease representation inequality, but also cause a big chunk of data to conceal in this way.

The total number of tweets belongs to each user helps to draw an overview of users' representation. However, this might be misinterpreted without consideration of temporal variation. The high representation of a user may indicate either over-representation regularly spread overtime or a specific situation that has greater importance for just one period of time. In other words, some of the users considered as over-represented themselves while they do this representation in a limited time interval but underrepresented for the rest of the time. In respect to that, representation varies among users, likewise, it varies for a user temporally due to several circumstances such as seasonal (summer or winter), emergencies (natural disasters, terror attacks), politics (election, referendum).

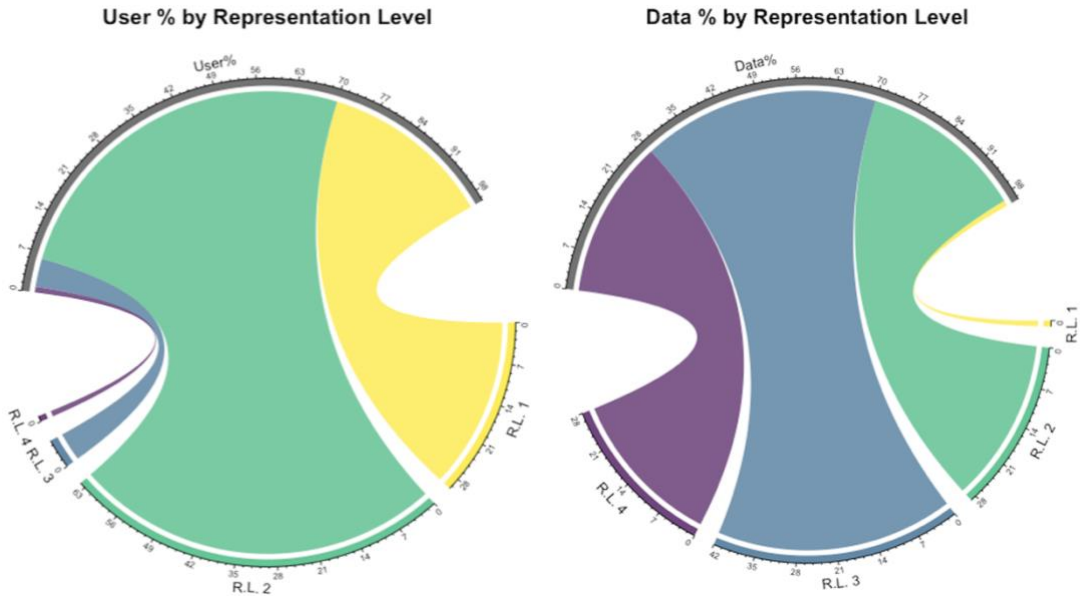
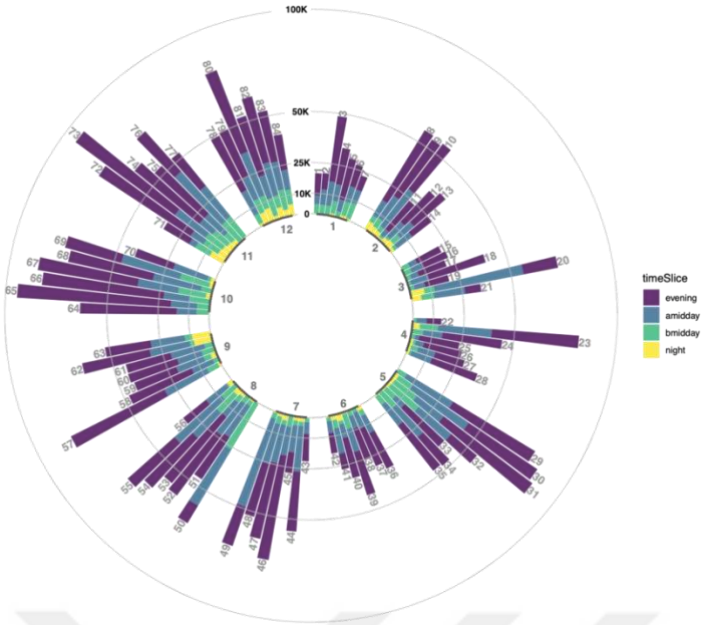


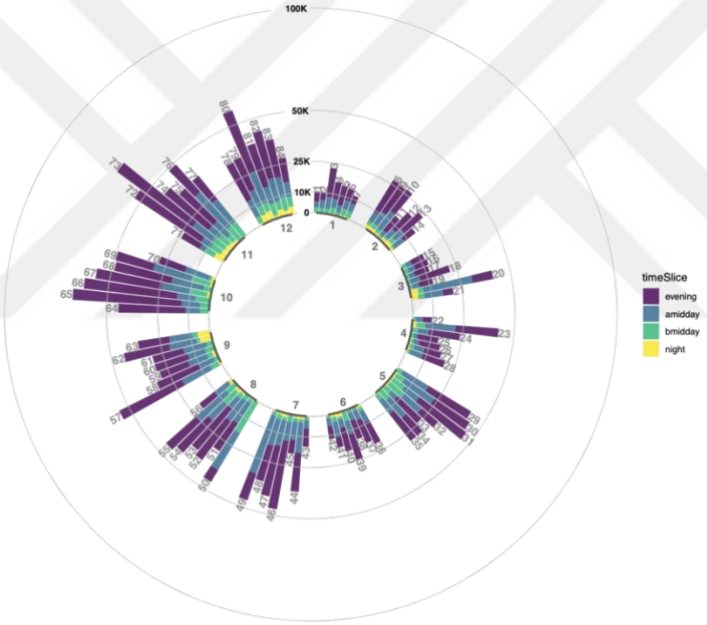
Figure 4.8 : Representation level (R.L.) by **(a)** percentage of users; **(b)** percentage of data.

In order to explore a number of tweets to each temporal level in one hand, a circular bar plot is combined. This gives the general temporal variation view of tidied data with some delusions due to inequalities in user/bot representations. Each bar in the plot represents daily data size according to the stacked time slices of the day as it is explained in the methodology section. Data without any weights considering user representation and normalization is visualized in Figure 4.9 (a). In respect to that, data production during a night time interval is pretty low or almost none for some days not because of system failure but because of temporal reasons. The biggest number of tweets is generated in the evening time for nearly every day though it is explicitly less than after middays for some days such as 20, 49, 50 (Figure 4.9 (a)).

There might be several misinterpretations while assessing the number of tweets in regards to time slices due to diverse representation levels of users. In order to avoid this, number of tweets to temporal level 1 is normalized by weight assignment to each user. According to that normalization, tweet counts for each time slice are recalculated by taking the sum of each user tweet count multiplied with its user weight. Normalized data is displayed in Figure 4.9 (b) as similar to Figure 4.9 (a). The number of tweets is obviously decreasing for all bars and a higher decreasing relative to previous numbers means higher overrepresentation level is normalized for time slices.



(a)

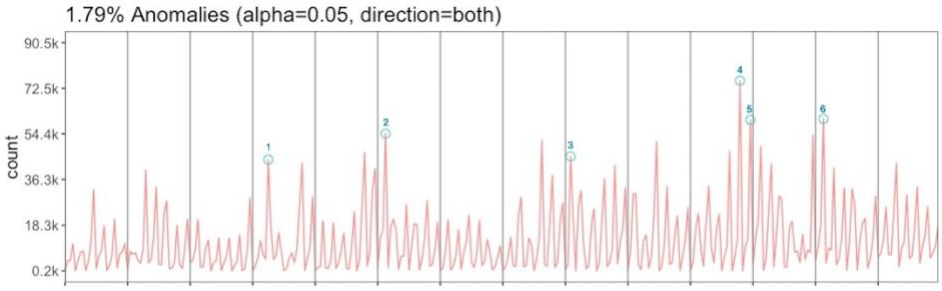


(b)

Figure 4.9 : Circular stacked bar plots for the 2018-year data **(a)** number of tweets in each temporal level; **(b)** normalized number of tweets in each temporal level.

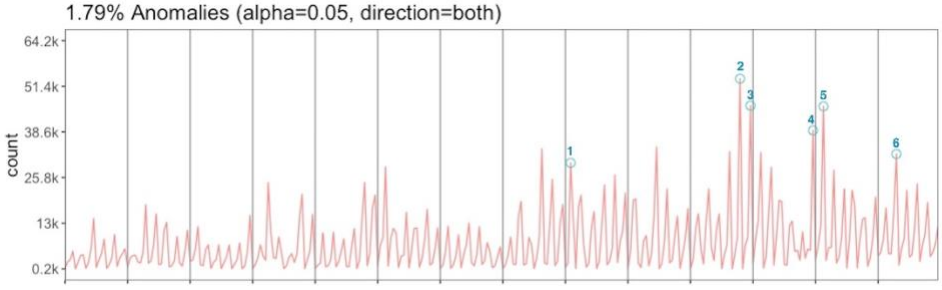
Diversity in representation level might create uncertainty on the number of data and requires to question whether data has any trend to do further inferences depending on that. In this general perspective without any location-based dimension, data is assessed with the anomaly detection algorithm in order to extract any trending activity. Three anomaly assessments are performed over tweet count (a), user count (b) and normalized tweet count (c) (Figure 4.10) respect to the time vector. While tweet count and user count have 6 anomaly slices which 4 of them are matched with each other,

the normalized count has 3 anomaly slices that are all the common with both tweet and user count anomalies. According to these matches, diversity in representation creates three more anomalies than the normalized representation. However, the matching anomaly slices 195 between the tweet and user count is also notable to be considered even it is not in the normalized count anomaly. Anomaly values in overall data are detected at 1.79%, 1.79%, and 0.89% respectively. This can be interpreted that data has strong trends for the assessment of 24 periodic time slices.



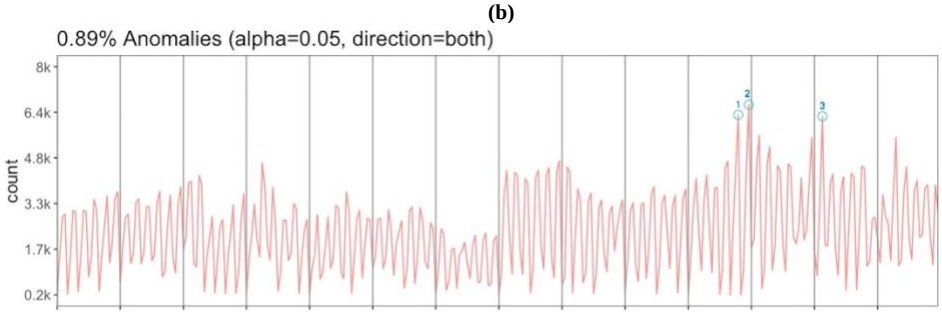
Label	Time Index	Anom Value
1	79	442502
2	124	545293
3	195	455144
4	260	754735
5	264	599536
6	292	60343

(a)



Legend Label	Time Index	Anom Value
1	195	299252
2	260	535513
3	264	459844
4	288	390525
5	292	45793
6	320	32429

(b)



Legend Label	Time Index	Anom Value
1	260	63172
2	264	66643
3	292	6268

(c)

Figure 4.10 : Anomaly in **(a)** number of tweets; **(b)** number of users; **(c)** normalized number of tweets.

Besides diverse representation levels of users, spatial representation is another aspect in order to understand data. Istanbul bounding box is divided into 100m x 100m grids to visualize this representation as represented and “unrepresented” grids as missing data. Figure and ground map (Figure 4.11 (a)) of the missing data perfectly match the

shape of the city (Figure 4.11 (b)). This matching can be taken as Twitter is a living bionic tool for Istanbul that more or less represents the living area of the city.

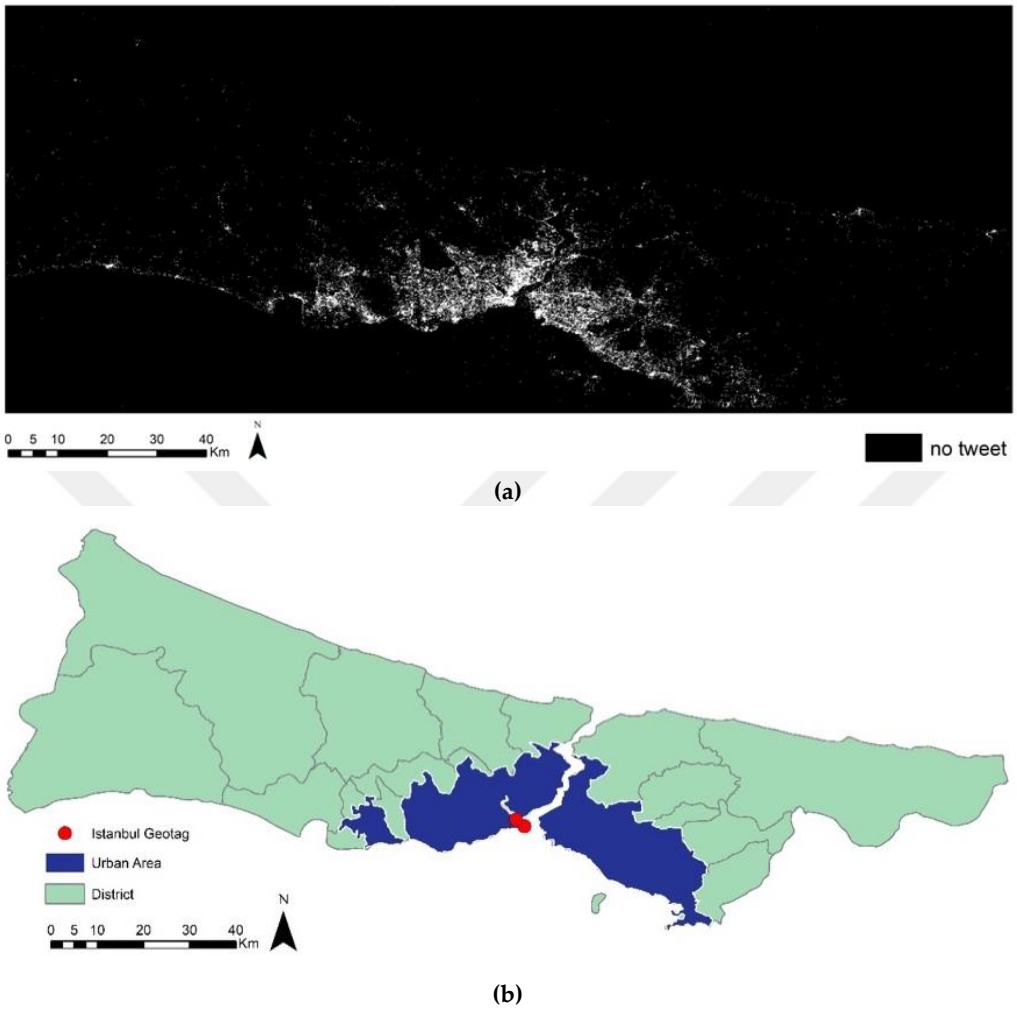


Figure 4.11 : Istanbul city **(a)** figure and ground map for tweet representations; **(b)** urban area and social media geotags.

A lattice-based monitoring system was designed in order to understand the spatial footprints of users. Grid size is determined as 1km x 1km since it is well enough for fine-grained event detection. The number of tweets is spatiotemporally normalized corresponding to the grids. In order to explore the spatial dynamics of the Istanbul, anomaly analysis is performed for each grid over the spatially normalized tweet count values as time vector with 336 slices. Each detected anomaly values for a grid is normalized with the overall anomaly amount detected in its time interval. In this way, the detected anomaly magnitude for a grid is calculated regarding the overall anomaly. The normalized anomaly rate per grid was calculated by adding all anomaly magnitudes for a grid and visualized in Figure 4.12. This reveals the locations where

most likely to have an anomaly in the city. In addition, this anomaly tendency map was tested with global and local Moran's I spatial correlation algorithm. The global I score and the p-value were found 0.24 and less than 0.0001 respectively which means the values were slightly positively correlated and the significance of the test is pretty high. In order to assess the positive and negative spatial autocorrelation, the anomaly tendency map was also tested with local Moran's I. It appears from Figure 4.12 (a), there are high anomaly rates in the center part of Istanbul, the local Moran's test confirms that there is positive spatial autocorrelation with the high positive Ii (Figure 4.12 (b)) and low p-value (Figure 4.12 (c)) in this area.

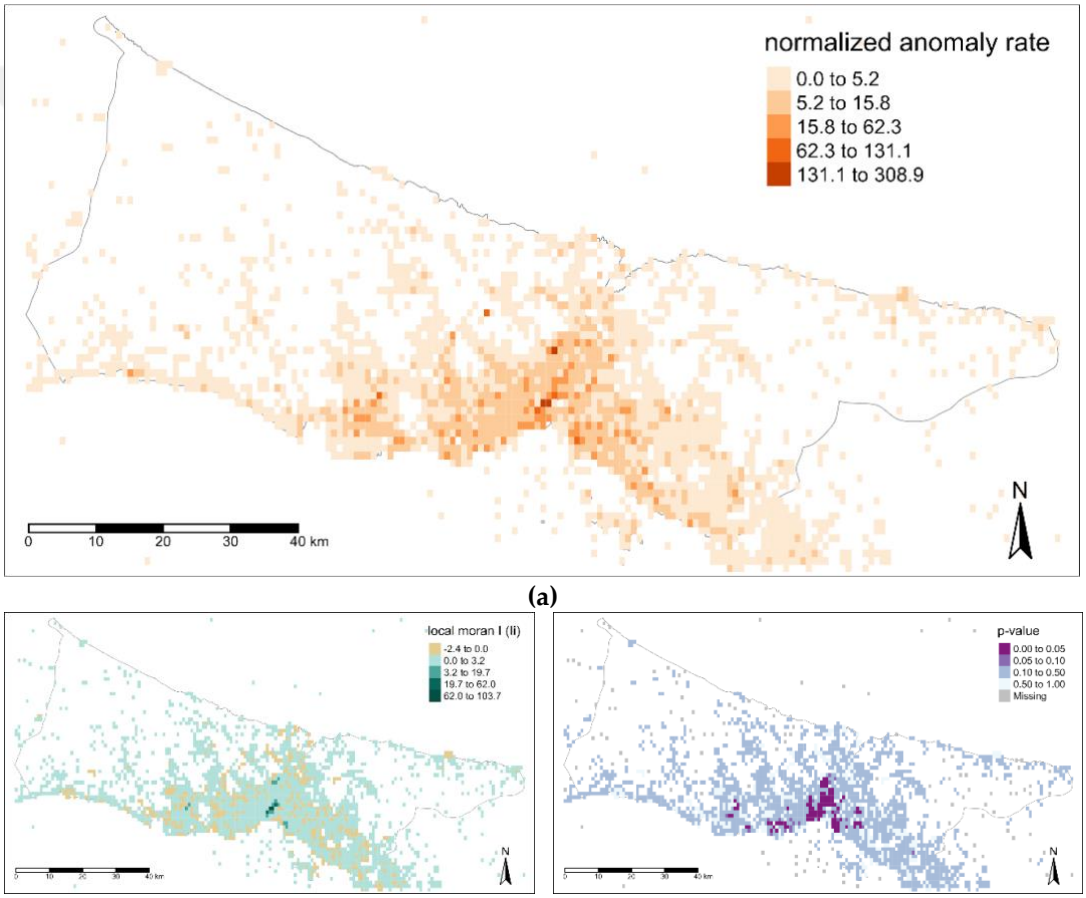


Figure 4.12 : Anomaly tendency assessment (a) Overall anomaly rate; (b) local Moran's I; (c) local Moran's p-value.

In order to understand general dynamics of the city, data was discretized from its anomalies detected previously. The anomalous values were replaced with the expected values for the related part of the data. In Figure 4.13, the average value of this retrieved trend data was represented in 4 classes. The first class includes the two most active

grids that have an average of 535 and 10473 for 6-hour. These spots are outliers and the closest value to the outliers is approximate to 50 tweets average in defined temporal level. There are two main reasons behind these outliers. The first, Istanbul geotags of social media platforms (Figure 4.11 (b)) are located within these grids. The second, spots are located in the central area of Istanbul (Figure 4.11 (b)) where the old city and touristic attractions are dense. The second and third classes grids by their activity have nearly the other 10% of the grids. The area covered with the second and third classes is matching with the urban area of Istanbul. These darker green grids show the central location bias of the data for general terms but also give the opportunity of monitoring them with the higher capacity of users' representation. The last class that covers nearly 90% of spatial grids include less than 1 tweet average within the 6-hour interval. This lowest active class comprehends mostly residential areas, rural parts, and seaside. While general activities within the darker areas presumed to be easily inferred from the tweet contents, the lowest active area could be easily spotted when there are extraordinary events.

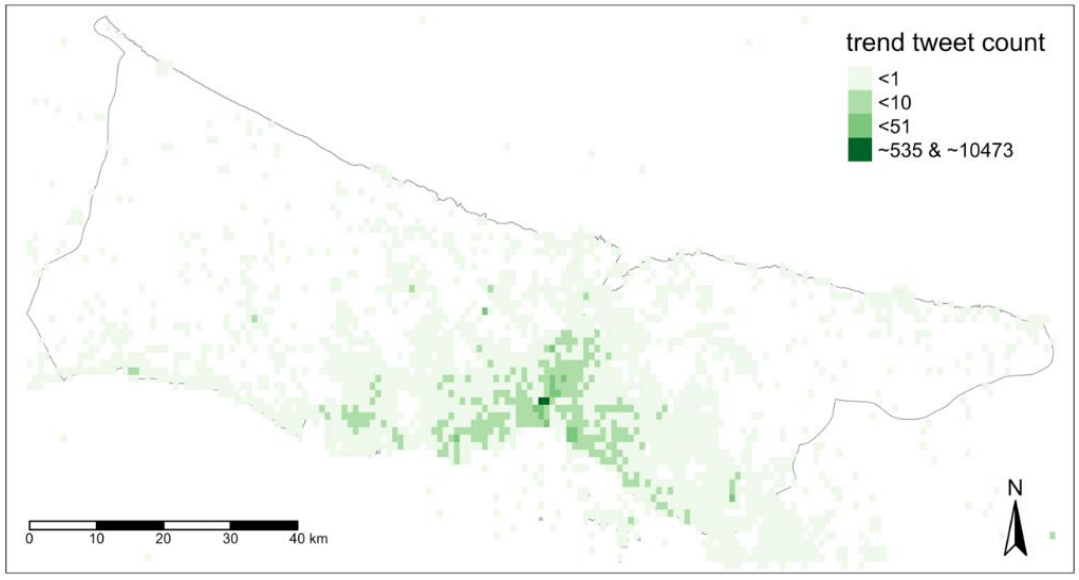


Figure 4.13 : Average tweet count of trend values in each grid within 6 hours.

Time is another aspect to detail spatial data review and citizens' footprints might vary in terms of different time levels. Therefore, the comparison between the trend map and maps belonging to different time levels was tackled at three temporal levels. The difference values defined with five classes that are below trend (least-low, low-trend), trend, above-trend (trend-high, high most). Difference values have positive and negative numbers, in addition, with the few exceptions values are not differentiated to

much in these maps. In respect to this, threshold values for these classes were determined manually after exploring data with several automatic classification techniques (such as; quantile, equal, standard deviation, kmeans, etc.). In Figure 4.14, the night map (a) has lower values than the trend map in the central parts while fringe parts of the urban areas have near values to trend. The difference in the second map (b) is more diverse and in some distributed areas the value is higher than the trend. In the third map for after midday time (c) and fourth map for evening timsse (d), the difference was reversed as the values are higher nearly in all parts of the city but dense in the central area of Istanbul (Figure 4.14 (a)).

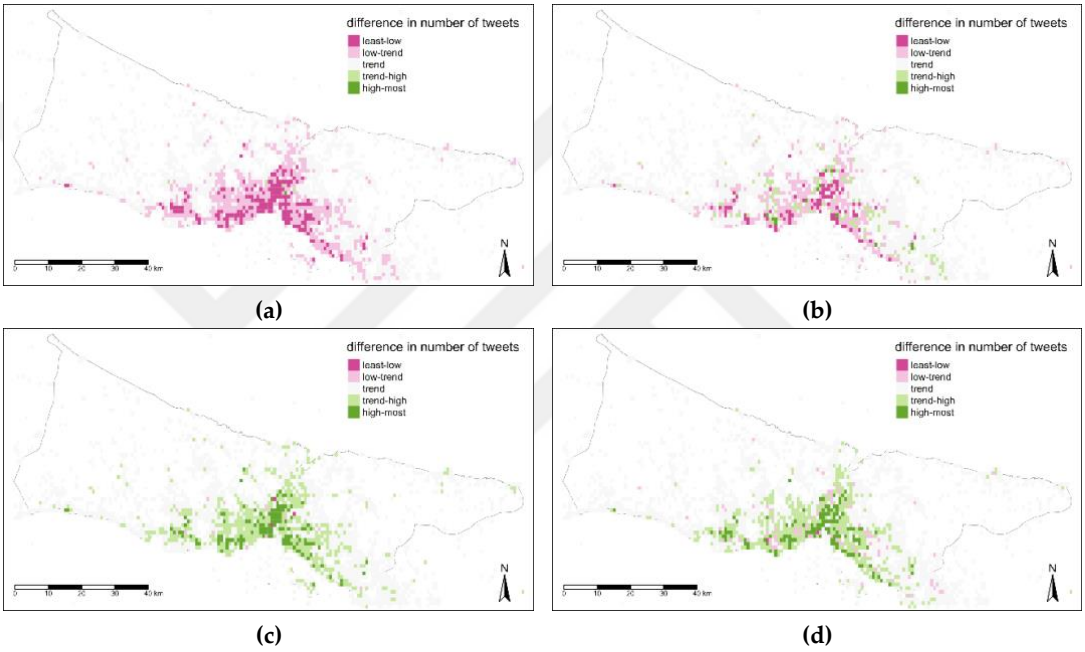


Figure 4.14 : Difference in the number of tweets between time level 1 and trend maps (a) night; (b) before midday; (c) after midday; (d) evening.

For the second temporal level, days of the week were considered. It is explicitly seen, maps of weekdays (a, b, c, d, e) are pretty similar to each other while the weekend days (f, g) apart from them with the spots having higher values near Istanbul Strait and along seaside (Figure 4.15). Basing on this weekday's map, there are no direct clustered values for a region and central areas are a mix of classes above and below trend values. The most part belongs to classes that are less than the trend in weekdays, while the most and specifically the central and coastline parts have higher values than trends for weekends.

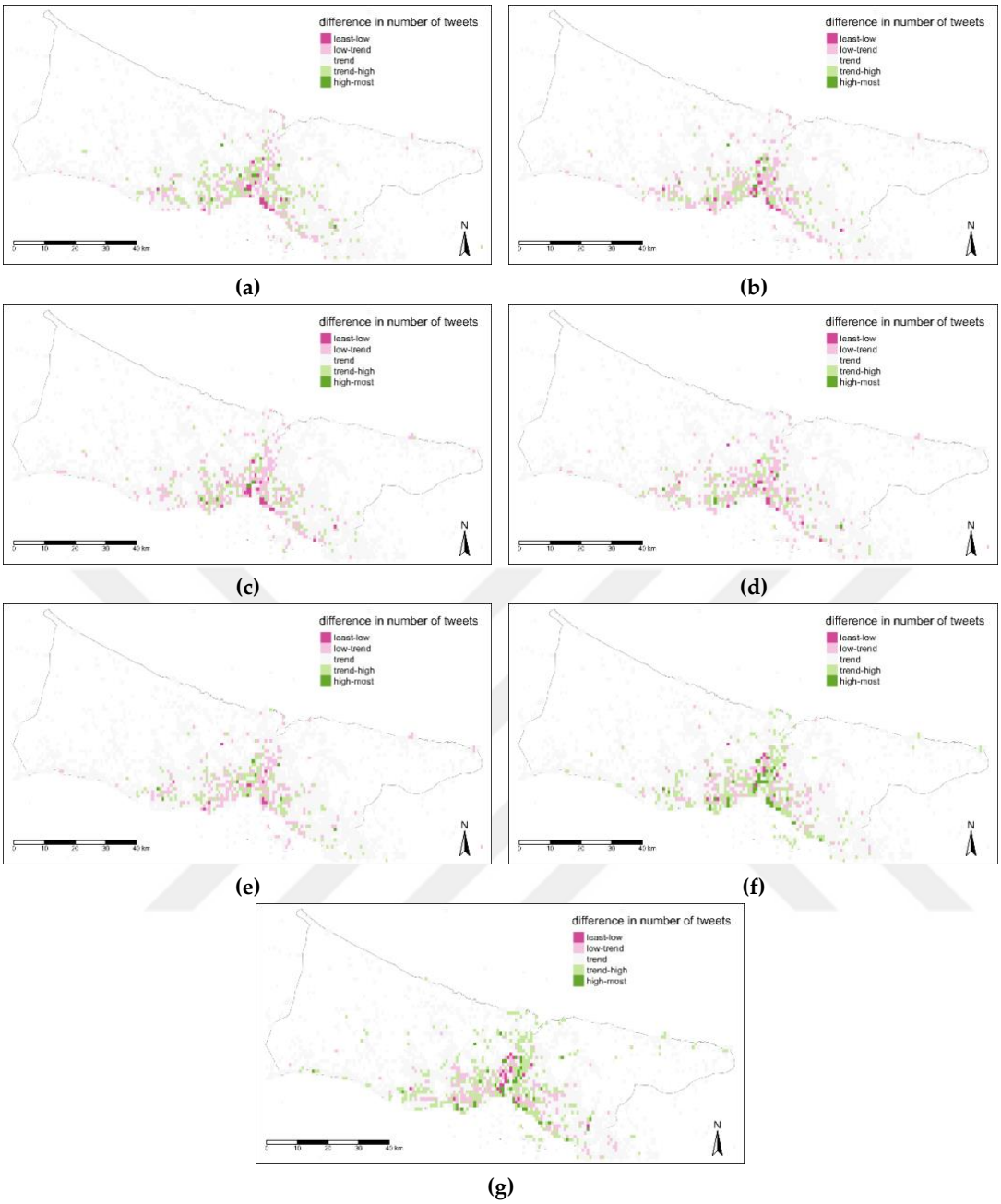


Figure 4.15 : Spatiotemporal Bias for Temporal Level- 2 (a) Monday; (b) Tuesday; (c) Wednesday; (d) Thursday; (e) Friday; (f) Saturday; (g) Sunday.

In the third time level assessment, seasons of the year were mapped (Figure 4.16). The central area has high value spots in winter and spring seasons while this area has less value in summer and autumn seasons (Figure 4.16). These seasonal plots reflect one side class for all parts of the city, that are above trend for winter and spring, below trend for summer and autumn. Global Moran's I was adopted to test the comparison maps' values spatial correlation. Though the I value is changing between -0.1 and 0.1 for each map, and the p-values are above 0.1, it is not possible to say the time variance is spatially auto correlated. That means there is no significant difference between the

difference values spatially although there is certain difference between the trend and time level maps as those seen in Figure 4.14, 4.15, and 4.16.

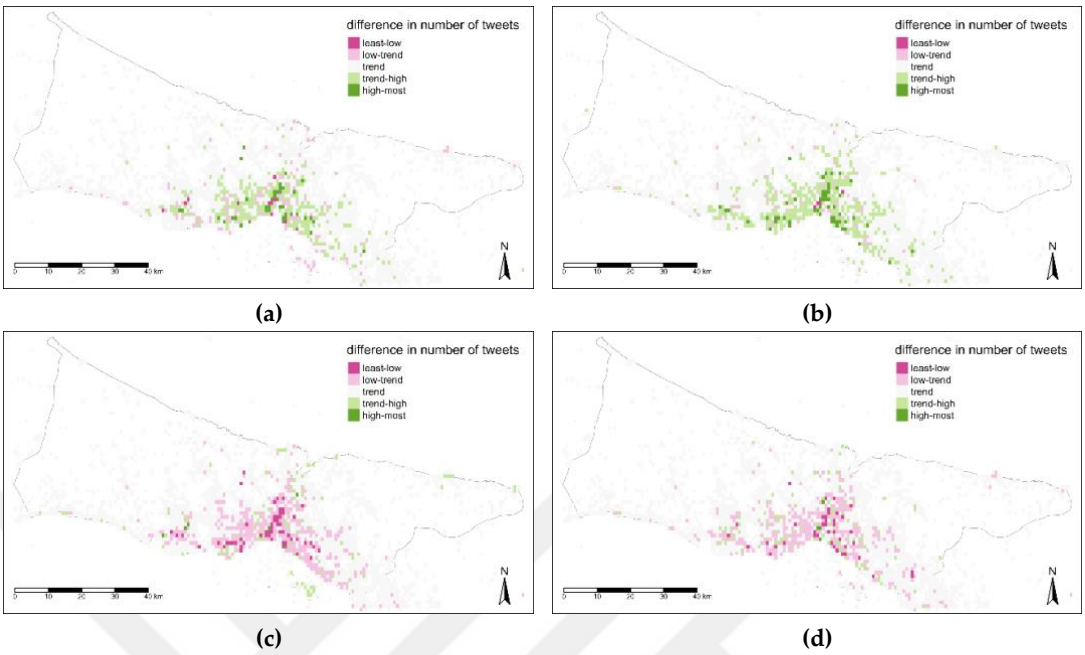


Figure 4.16 : Spatiotemporal Bias for Temporal Level- 3 **(a)** Winter; **(b)** Spring; **(c)** Summer; **(d)** Autumn.

4.5 Discussion

Social media is an invaluable source of data that is generated by human sensors due to its immense sensing capability and continuity (Goodchild, 2007a; Zhao et al, 2011). Though it has various type of accountholders (Zi et al, 2010), the content and the spatial activity of each of them varies too (Cheng et al, 2010). There are researches providing background information to explain the activities of users on social media and categorizing them (Issa et al, 2017; Sloan and Morgan, 2015). And some researches on the credibility of users (Castillo et al, 2013; Wang and Zhuang, 2018) and some others try to determine coordinated users who behave together and manipulate data content (Abbasi and Liu, 2013). And studies claims social media is full of rumors and most part of the accountholders spreading wrong information while there are emergencies and do not correct the content even if they are informed later (Ma et al, 2018; Middleton and Krivcovs, 2016). In respect to this, social media data requires to be assessed without removing any data but accepting all these deficits and considering them with the nature of itself, since it is not possible to control each user credibility in real-time without historical data or demographic information. Although data includes several issues such as credibility, rumors, representation inequalities in

terms of user, it has a reference pattern in order to be assessed for monitoring systems. This study evaluated and presented general citizen footprint, most likely regional anomaly maps and spatiotemporal biases in Istanbul. These inferences as reference maps provide interpretation easiness for monitoring the city.

This study evaluates a-year SMD with the methodology given in subsection 4.3.1. From the data revealed in this study it has been realized that to investigate the spatiotemporal change in SMD, the difference in representation levels of users should be normalized. For spatial exploration, miscellaneous representations of users are avoided with the spatio-temporal normalization technique. Anomalies that data might have due to an unusual event or coordinated users' activity, detected and replaced with the expected value. The locations, which tend to have more anomaly counts are determined as location biased spots. And the data norm displays the most represented locations by the more account holders. Besides that, the data norm is used as reference to explore spatiotemporal biases. It is obvious that data has several kinds of biases and the evaluated data can be used as a reference to discriminate any abnormality.

The results of this study can be enhanced with the finer spatial grain for lattice-based monitoring and finer time grain instead of 6-hour. Further studies can also be developed over data content by doing several text analyses in order to find the most co-occurred word in a lattice and make a contribution to the reference map in that way.

Istanbul is the most populated city in Turkey with over 15 million citizens (Turkish Statistical Institute, 2019) and 3 million visitors which makes this city very important to be monitored for the sake of the living standards and responsive emergency management as well. There are several smart city projects separate conducted by local authorities; however, those projects are limited with the base map digitization or some municipality paperwork processes. Since citizens of Turkey have high potentials to generate spatial data on Twitter compared to other many countries, Twitter is eligible for citizen-based projects in Istanbul (Sloan and Morgan, 2015). In further studies, evaluated data in this study provide the benchmarking knowledge to establish a dynamic monitoring system for Istanbul.

This study exposed four outcomes as mentioned below. The first outcome reveals that highly active users generate the majority of the data and as a general approach, removing this data within a pseudo-cleaning process conceals a large amount of data.

The second one is the anomaly outcome results changes due to the diverse representation levels of the users. That is why; data normalization in terms of representation levels plays an important role in the detection of the true anomaly. The third outcome exposes that, as shown in Figure 4.12 (a), spatiotemporally normalized data represent strong spatial anomaly tendency at the urban center. The last outcome shows that the trend data is dense in the urban center and the spatiotemporal bias assessments show the data density varies in terms of the time of day, day of week and season of the year.

Twitter API is used in this study as in commonly used for other academic studies. Twitter declares that this API provides randomized 1% of public tweets in real-time (Twitter, 2020b). There are empirical researches that test this randomization by comparing this sampled amount with the Firehose API data which provides the whole public tweets (Morstatter et al, 2013, 2014). Studies found out there is no significant indication that the sampling of the Twitter API is biased with one exception since Twitter randomized the tweets by assigning ID for each tweet regarding the millisecond time. Because, this randomization is plausible since it exceeds a person's capability to share tweets in that quickness, unlike a bot. The used data within this study is normalized both spatially and non-spatially to avoid representation bias that can also eliminate noise due to bot accounts.

In this study, PostgreSQL with the PostGIS extension is used for data handling. PostgreSQL is an open-source relational database management system (RDBMS) that can be deployed to different environments such as a desktop, a cloud or a hybrid environment database. The storing capacity and the time cost for the processes rely on the specifications of the environment. This relational database is adequate to handle a large amount of data for basic operations (such as; insert, select and update) as in this study but could not be the best option for big data studies transactions (Jung et al, 2015). NoSQL database like MongoDB has enhanced functionality on the big data processing performance especially the ones performed on unstructured data. SMD has unstructured content and RDBMS has issues while structuring the big amount of data. For this reason, NoSQL should be preferred while processing the big amount of unstructured text of SMD (Jung et al, 2015; Mathew and Madhu Kumar, 2015). This work is also planned to extend with other cities' data including text mining in the

context of space-time and the finer-grained temporal analyses. Therefore, in further studies a NoSQL database management system will be considered to handle such data.

There are numerous studies in order to conceptualize the measures of data quality in the context of VGI (Ballatore and Zipf, 2015; Senaratne et al, 2017). Data quality studies on VGI are mostly tackling data quality measures such as; completeness, positional accuracy, and granularity on Open Street Map (Ballatore and Zipf, 2015; Haklay, 2010; Mocnik et al, 2018). Generally, approaches for data quality are assessed in two class as intrinsic and extrinsic. In the intrinsic assessment, there is no use of an external reference map unlike extrinsic data quality assessment (Mocnik et al, 2018). SMD was assessed in several aspects in this study in order to understand data bias, anomalies and trends. Since there is no external data use for these assessments within this study, this study provides the methodology for assessing the intrinsic data quality of the SMD.

The methodology proposed in this study can be used to extract the unbiased daily routines of the social media data of the regions for the normal days, and this can be referred for the emergency or unexpected event cases to detect the change or impacts. Data assessment in this study is based on revealing the citizen footprints in SM and designed to explore anomalies, trends, and bias within data.

In further studies, inferences from this study will be used to functioning a citizen-based monitoring system for Istanbul. The system design conceptually will follow the steps; tweets are collected in real-time, a number of tweets from the distinct users for each spatial grid is calculated, normalized number of tweets is assessed with the anomaly detection algorithm with regression line over trend data, detected anomalies are assessed with the most likely regional anomaly maps and the decision is made for emergency conditions. In addition, the proposed methodology is planned to work out on other big cities in order to contribute to other researchers by providing the results (reference maps) on a designed webpage for our future projects.

5. CONCLUSIONS AND RECOMMENDATIONS

Developments in communication technologies let the emergence of various data sources in the last decade. SM is one of them that provides a wide variety of data from human-based sensors. Unlike generic sensors, human-based sensors provide unstandardized data. Therefore, content variety and accuracy of SMD depend on the account holders. In addition, sensors have neither a certain temporal nor spatial resolution for providing data. This means uncertainty of providing data for a researched topic in a determined spatiotemporal resolution. Since SM is invaluable data source for disaster management, this thesis focused on these uncertain issues of SMD to make this value more understandable and interpretable.

The obstacle to SMD use is the uncertainty in terms of variety, capacity, and accuracy of data. Accountholders with different backgrounds and interacting with different places reflect who and where they are (Cheng et al, 2010; Sloan and Morgan, 2015; Zi et al, 2010). In respect to this, each region has its own SMD nature that is formed by its own citizens or visitors. In consideration with that, Istanbul administrative boundary is chosen as the case study area for each paper published as a part of this thesis.

The first study is based on determining the capacity of SMD to spatially detect a man-made disaster. SMD was collected along a day from the beginning of the coup attempt in Istanbul, Turkey, and processed as an hourly hot spot maps in order to discover the capacity of SMD to detect incidences in a fine-grained spatial grid for each hour. The produced hotspot maps for each hour were exploratively checked with the incidence locations that are extracted from newspaper sources and found correlated. Lastly, results were compared and tested with the dataset belongs to one week earlier whether this correlation was a coincidence due to spatial data bias. According to that test, it was revealed that the results are not caused by data bias. Consequently, the paper reached its aim to understand the data capacity on detecting incidences in Istanbul.

The approach that was used in the first study simply depends on mapping data with spatial hot spot analysis. However, there are several steps that can upgrade the results

for fine-grained spatial information extraction with the content classification. Filtering the relevant content is highly important to define the incidence location when there are several hot topics. SMD should be filtered in terms of a domain in order to retrieve directly relevant content. Since the data content is unstructured, filtering SMD is an issue that depends on reliability of the chosen NLP techniques. For this reason, in the second paper, SMD was assessed in the perspective of fine-grained domain-based filtering and its impacts on spatial incidence mapping.

In the second paper, SMD related to two terror attacks happened in Istanbul were used for the case study. Since Turkish is commonly used language in Turkey so as within the collected data, domain-based filtering techniques was adapted to Turkish languages. Several techniques based on subjectivity lexicons and ML were utilized to filter relevant data and each result was assessed with the statistical scores. This was the first part of the reliability of SMD process related with the text filtering. The second part was finding the impacts of this filtering process on spatial reliability. For this reason, spatial reliability of incidence maps, produced with the filtered data by each filtering methodology, was assessed in order to understand how the filtering might affect the results. This was a process that requires a quantitative measure instead exploratively comparing two maps. Because interpreting the results of exploratory analyses, take time and might be hard for wide-area analysis having lots of incidents as a result. In order to measure the spatial reliability, it was decided to use the spatial similarity indices as a quantitative method. This spatial reliability calculation based on the usage of spatial similarity indices by comparing the incidence maps produced with the use of manually filtered and automatically filtered data. At this point, a new spatial similarity index (Giz Index) was formulated specifically for the use of incidence mapping comparisons since the currents disregard the proximity of non-intersected clusters or incidences in either prediction or ground truth dataset. But the close things can be related with each other and especially in incidence mapping this can not be disregarded according to Tobler's First Law of Geography. Giz Index regards the proximity by weigthing it to the size of the related cluster area. Giz Index with this feature provides a fast and quantitative reliability control for the spatial outcomes of the SMD studies. By this way, second paper contributes to the reliability search of SMD analyzing techniques and methodologies in a quantitative way.

As well as the used techniques in SMD analyses, data quality is another aspect that affect the result accuracy. In the third paper, SMD dynamics in Istanbul were investigated in order to understand intrinsic data qualities such as; anomalies, trends and bias. The study presents a methodology for SMD investigation with the several inferences for the data of Istanbul. The first outcome of the study, the majority of SMD were generated by unproportionally less amount of accountholders. In other words, SMD has representation bias that should be normalized in order to avoid data manipulation by a small group of account holders. The other inference over representation is temporal effect on data amount. Temporal impacts on data were assessed in terms of three-time levels as time of day, day of week and season of year. According to that, night time data amount is seen gradually lower than all the other time intervals of the day while the evening time data amount is higher than. For the other time intervals assessment, it is uncertain to say the amount varies in terms of the day of week and the season of year. In addition, SMD is assessed spatially after data normalization and tidying processes. First spatial investigation of SMD was the spatial representation. Although some parts of the Istanbul area are spatially unrepresented or very low represented, spatially represented parts refer to the urban areas. This can infer that the use of SMD for monitoring Istanbul is eligible since there are citizens' footprints in the most part of the urban area. The second investigation was the anomaly assessment and the study found out some parts of the Istanbul have more tendency to be anomalous apart from others. Therefore, these areas should be considered carefully since these areas can be assumed more responsive than the other urban parts. The third investigation was the determination of spatial trends of SMD. It was discovered that some parts have high spatial representation rate than others due to default geotagging locations for Istanbul City. And this trend map is used as a reference map for the fourth investiagation which included a series of spatio-temporal bias in Istanbul. All these steps that was designed for the data investigation are the base of a city monitoring system with the SMD. By this way, the third paper of this thesis aims to contribute to the methodology of monitoring a city with SMD and interpret the results with the produced reference maps (anomaly, trend and bias maps).

This thesis was composed of two published articles and one book chapter. Regarding the aim of this thesis, each paper evaluates the SMD in different perspectives of information mining methodologies which are data capacity, SMD mining

methodologies, and techniques reliability, and data intrinsic quality investigation. At first, the studies show that basically SMD has the capacity to be used for spatial information extraction as coarse grained mapping activities. However, for a fine-grained information extraction, both text mining methodologies and spatial assessment techniques should be considered quantitatively. In addition to that, the success of the filtering methods used for fine-grained incidence mapping are only enough for the spatial information accuracy. In order to reach the high accuracy in incidence mapping, data dynamics should be considered in terms of data biases.

All the attempts in this thesis were for increasing the accuracy of fine-grained incidence mapping with the use of SMD and found useful regarding the performance of the case study results. The credibility of the data source is another aspect that may also affect the results of information extraction and assumed the misinformation is avoided due to the easing affect of high data amount as multiple-check. On the other hand, determination of credibility for each post may highly affect the accuracy of the resulted maps when and where there is data scarcity. Since the given information in the text content, geotags can also be manipulated by the users. The initial understanding for that manipulation can be obtained from the 2.3.1 Data accountability. According to that a social media user can geotag or georeferenced the post somewhere else than the current location. This location can be even very far from the exact location of the user. Furthermore, the content may not be related with the geotagged location. In other words, the content and the location of the post can be inconsistent. Determining this inconsistency for each post is not possible since there is no location text for geoparsing of each content. In respect to that, the determination of spatial credibility of users is another comprehensive study topic that is planned as future works.

REFERENCES

Abbasi, M. A., and Liu, H. (2013). Measuring user credibility in social media. *In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, 441-448. Springer, Berlin, Heidelberg.

Acar, A., and Muraki, Y. (2011). Twitter for crisis communication: Lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, *7*(3), 392–402.

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., and Liu, B. (2011). Predicting Flu Trends using Twitter data. *In 2011 IEEE Conference on Computer Communications Workshops*. IEEE.

Aksoy, A., and Ozturk, T. (2018). Turkish Stop Words. Retrieved May 1, 2016. Available from https://github.com/ahmetax/trstop/

Alexander, D. (1993). Natural Disasters Chapman & Hall. *Inc*. 632.

Anderson, L. T. (1995). Guidelines for preparing urban plans.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115.

Ao, J., Zhang, P., and Cao, Y. (2014). Estimating the locations of emergency events from Twitter streams. *Procedia Computer Science*, *31*, 731–739.

Aramaki, E. (2011). Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter The University of Tokyo The University of Tokyo National Institute of. *Computational Linguistics*.

Arnesson, A., and Lewenhagen, K. (2018). *Comparison and Prediction of Temporal Hotspot Maps* (Master's thesis). Available from http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1228347&dswid=8431

Arthur, R., and Williams, H. T. P. (2019). The human geography of Twitter: Quantifying regional identity and inter-region communication in England and Wales. *PloS One*, *14*(4).

Aytekin, Ç. (2013). An opinion mining task in Turkish language: A model for assigning opinions in Turkish blogs to the polarities. *Journalism and Mass Communication*, *3*(3), 179–198.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, LREC 2010.

Ball, J. (2002). Towards a methodology for mapping 'regions for sustainability'using PPGIS. *Progress in Planning*, *58*(2), 81–140.

Ballatore, A., and Jokar Arsanjani, J. (2018). Placing Wikimapia: an exploratory analysis. *International Journal of Geographical Information Science*, *33*(8), 1633–1650.

Ballatore, A., and Zipf, A. (2015). A Conceptual Quality Framework for Volunteered Geographic Information. *In Spatial Information Theory*, 89–107. Springer International Publishing.

Basiri, A., Haklay, M., and Gardner, Z. (2018). The impact of biases in the crowdsourced trajectories on the output of data mining processes. *Association of Geographic Information Laboratories in Europe (AGILE)*.

Basiri, A., Haklay, M., Foody, G., and Mooney, P. (2019). Crowdsourced geospatial data quality: challenges and future directions. *International Journal of Geographical Information Science*, *33*(8), 1588–1593.

Bassolas, A., Lenormand, M., Tugores, A., Gonçalves, B., and Ramasco, J. J. (2016). Erratum to: Touristic site attractiveness seen through Twitter. *EPJ Data Science*, *5*(1).

Bekkar, M., Djemaa, H. K., and Alitouche, T. A. (2013). Evaluation Measures for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, *3*(10).

Berren, M. R., Beigel, A., and Ghertner, S. (1980). A typology for the classification of disasters. *Community Mental Health Journal*, *16*(2), 103–111.

Bilham, R. (2010). Lessons from the Haiti earthquake. *Nature*, *463*(7283), 878–879.

Birch, C. P. D., Oom, S. P., and Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, *206*(3–4), 347–359. doi:10.1016/j.ecolmodel.2007.03.041.

Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., and Blanchet, G. (2015). Package 'spdep.' Retrieved December 9, 2015. Available from http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/packages/spdep.pdf.

Boccardo, P. (2013). New perspectives in emergency mapping. *European Journal of Remote Sensing*, *46*(1), 571–582. doi:10.5721/eujrs20134633.

Bouchet-Valet, M. (2015). Snowball stemmers based on the C libstemmer UTF-8 library. *Comprehensive R Archive Network*.

Branco, P., Torgo, L., and Ribeiro, R. P. (2016). A Survey of Predictive Modeling on Imbalanced Domains. *ACM Computing Surveys*, *49*(2), 1–50. doi:10.1145/2907070.

Brooker, P., Barnett, J., and Cribbin, T. (2016). Doing social media analytics. *Big Data and Society*. doi:10.1177/2053951716658060.

Brown, G. (2017). A Review of Sampling Effects and Response Bias in Internet Participatory Mapping (PPGIS/PGIS/VGI). *In Transactions in GIS*. doi:10.1111/tgis.12207.

Bruns, A., and Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, *17*(4).

Bucklin, D., and Basille, M. (2018). rpostgis: Linking R with a PostGIS Spatial Database. *The R Journal*, *10*(1), 251. doi:10.32614/rj-2018-025.

Cakir, M. U., and Guldamlasioglu, S. (2016). Text Mining Analysis in Turkish Language Using Big Data Tools. *In 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*. IEEE. doi:10.1109/compsac.2016.203.

Cambria, E., Hussain, A., Durrani, T., and Zhang, J. (2012). Towards a Chinese Common and Common Sense Knowledge Base for Sentiment Analysis. *In Advanced Research in Applied Artificial Intelligence*, 437–446. Springer Berlin Heidelberg. doi:10.1007/978-3-642-31087-4_46.

Carto. (2014). CartoDB Twitter Maps: the definitive way to perform geospatial analysis and visualization of Twitter activity. Available from https://carto.com/blog/twitter-maps/

Castillo, C., Mendoza, M., and Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research*. doi:10.1108/IntR-05-2012-0095.

Ceron, A., Curini, L., Iacus, S. M., and Porro, G. (2013). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, *16*(2), 340–358. doi:10.1177/1461444813480466.

Cheng, Z., Caverlee, J., and Lee, K. (2010). You are where you tweet. *In Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10. ACM Press*. doi:10.1145/1871437.1871535.

Chen, J., Liu, Y., and Zou, M. (2016). Home location profiling for users in social media. *Information and Management*. doi:10.1016/j.im.2015.09.008.

Chen, X., Sin, S.-C. J., Theng, Y.-L., and Lee, C. S. (2015). Why Do Social Media Users Share Misinformation? *In Proceedings of the 15th ACM/IEEE-CE on Joint Conference on Digital Libraries - JCDL '15. ACM Press*. doi:10.1145/2756406.2756941.

Cieliebak, M., Deriu, J. M., Egger, D., and Uzdilli, F. (2017). A Twitter Corpus and Benchmark Resources for German Sentiment Analysis. *In Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. Association for Computational Linguistics. doi:10.18653/v1/w17-1106.

Cilden, E. K. (2007). SnowballC Turkish Stemmer. Retrieved May 1, 2015. Available from http://snowball.tartarus.org/.

Clarke, K. C. (1997). Getting started with GIS. Prentice Hall Upper Saddle River, NJ.

Conway, J., Eddelbuettel, D., Nishiyama, T., Prayaga, S. K., and Tiffin, N. (2016). RPostgreSQL: R interface to the PostgreSQL database system. *In CRAN*.

CRED. (2016). EM-DAT The International Disaster Database. Retrieved May 1, 2015. Available from http://www.emdat.be/.

Crooks, A., Croitoru, A., Stefanidis, A., and Radzikowski, J. (2013). #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS*. doi:10.1111/j.1467-9671.2012.01359.x.

Cutter, S. L. (2003). GI Science, Disasters, and Emergency Management. *Transactions in GIS*, *7*(4), 439–446. doi:10.1111/1467-9671.00157.

da Silva Meyer, A., Garcia, A. A. F., Pereira de Souza, A., and Lopes de Souza, C. (2004). Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (Zea mays L). *Genetics and Molecular Biology*, *27*(1), 83–91. doi:10.1590/s1415-47572004000100014.

D'Andrea, E., Ducange, P., Lazzerini, B., and Marcelloni, F. (2015). Real-time detection of traffic from twitter stream analysis. *Intelligent Transportation Systems*, IEEE Transactions On, *16*(4), 2269–2283.

Danielsson, P.E. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing*, *14*(3), 227–248. doi:10.1016/0146-664x(80)90054-4.

Dehkharghani, R., Saygin, Y., Yanikoglu, B., and Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*. doi:10.1007/s10579-015-9307-6.

Deshwal, A., and Sharma, S. K. (2016). Twitter sentiment analysis using various classification algorithms. *In 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. IEEE. doi:10.1109/icrito.2016.7784960.

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302. doi:10.2307/1932409.

Dong, Y.-S., and Han, K.-S. (2004). A comparison of several ensemble methods for text categorization. doi:10.1109/scc.2004.1358033.

Elnashai, A. S., Hampton, S., Karaman, H., Lee, J. S., Mclaren, T., Myers, J., …, and Tolbert, N. (2008). Overview and Applications of Maeviz-Hazturk 2007. *Journal of Earthquake Engineering*, *12*(sup2), 100–108. doi:10.1080/13632460802013750.

Elwood, S., Goodchild, M. F., and Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, *102*(3), 571–590. doi:10.1080/00045608.2011.595657.

EM-DAT. (2009). General Classification. Retrieved May 1, 2015. Available from http://www.emdat.be/classification.

Erdik, M., Cagnan, Z., Zulfikar, C., Sesetyan, K., Demircioglu, M. B., Durukal, E., and Kariptas, C. (2008). DEVELOPMENT OF RAPID EARTHQUAKE LOSS ASSESSMENT METHODOLOGIES for EURO-MED REGION. *The 14th World Conference on Earthquake Engineering*, October 12-17, 2008 Beijing, China.

Erdik, M., Şeşetyan, K., Demircioğlu, M. B., Hancilar, U., and Zülfikar, C. (2011). Rapid earthquake loss assessment after damaging earthquakes. *Soil Dynamics and Earthquake Engineering*. doi:10.1016/j.soildyn.2010.03.009.

Erogul, U. (2009). *Sentiment Analysis in Turkish* (Master's thesis). Available from https://open.metu.edu.tr/handle/11511/18638.

Eshghi, K., and Larson, R. C. (2008). Disasters: Lessons from the past 105 years. *Disaster Prevention and Management: An International Journal.* doi:10.1108/09653560810855883.

ESRI. (2018). Hot Spot Analysis (Getis-Ord Gi*). Retrieved May 1, 2020. Available from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm

ESRI. (2018). Optimized Hot Spot Analysis. Retrieved May 1, 2020. Available from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm

Feinerer, I. (2017). Introduction to the tm Package Text Mining in R. Available from https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf

Feinerer, I., Buchta, C., Geiger, W., Rauch, J., Mair, P., and Hornik, K. (2013). The textcat package for n-gram based text categorization in R. *Journal of Statistical Software*, *52*(6), 1–17.

Fellows, I., Fellows, M. I., Rcpp, L., and Rcpp, L. (2018). Package 'wordcloud.'

FEMA. (2000). PARTNERSHIPS IN PREPAREDNESS. Retrieved May 1, 2015. Available from https://www.fema.gov/media-library-data/20130726-1447-20490-2091/partners_v4.pdf

FEMA. (2007). Disaster Planning Is Up To You.

Feng, C.-C., and Flewelling, D. M. (2004). Assessment of semantic similarity between land use/land cover classification systems. *Computers, Environment and Urban Systems*, *28*(3), 229–246. doi:10.1016/s0198-9715(03)00020-6

Fonte, C. C., Bastin, L., Foody, G., Kellenberger, T., Kerle, N., Mooney, P., …, and See, L. (2015). VGI quality control. *ISPRS Geospatial Week 2015*, 317–324.

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* (Oxford, England), *31*(22), 3718–3720. doi:10.1093/bioinformatics/btv428

Gao, H., Barbier, G., and Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, *26*(3), 10–14. doi:10.1109/MIS.2011.52

Gardner, Z., and Mooney, P. (2018). Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study. *Proceedings of Agile Conference*. Lund: AGILE, 12–15.

Gelernter, J., and Balaji, S. (2013). An algorithm for local geoparsing of microtext. *GeoInformatica*, *17*(4), 635–667.

Gelernter, J., and Mushegian, N. (2011). Geo-parsing Messages from Microtext. *Transactions in GIS*, *15*(6), 753–773.

Gelernter, J., and Wu, G. (2012). High performance mining of social media data. *Proceedings of the 1st Conference of the Extreme Science and*

Engineering Discovery Environment: Bridging from the EXtreme to the Campus and Beyond, 22.

Gengec, N. E. (2016). Geo Tweets Downloader. Retrieved May 1, 2016. Available from https://github.com/nagellette/geo-tweet-downloader/

Gentry, J. (2012). Twitter client for R.

Getis, A., and Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. *In Perspectives on spatial data analysis*, 127–145. Springer, Berlin, Heidelberg.

Go, A., Bhayani, R., and Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *CS224N project report*, Stanford, *1*(12), 2009.

Goodchild, M. F. (2007a). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221.

Goodchild, M. F. (2007b). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, *2*(2), 24-32.

Goodchild, M. F. (2009). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, *3*(2), 82–96.

Green, W. G., and Richmond, V. (2003). CHANGES TO THE DISASTER DATABASE PROJECT.

Gross, D., and Hanna, J. (2010). Facebook introduces check-in feature. *CNN*. http://edition.cnn.com/2010/TECH/social.media/08/18/facebook.location/index.html

Gulnerman, A. G. (2020). tr_text_clean. Retrieved January 1, 2020. Available from https://github.com/gulnerman/tr_text_clean

Gulnerman, A. G., Gengec, N. E., and Karaman, H. (2016). Review of Public Tweets over Turkey within a pre-Determined Time. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *3*(1), 153–159.

Gulnerman, A. G., Goksel, C., and Tezer, A. (2017). Disaster Capacity Building with a GIS Tool of Public Participation. *Fresenius Enviromental Bulletin*, 26(1), 237–243.

Gulnerman, A. G., and Karaman, H. (2015). PPGIS Case Studies Comparison and Future Questioning. *In 2015 15th International Conference on Computational Science and Its Applications*, Banf, Canada, 104 – 107. IEEE.

Gulnerman, A. G., and Karaman, H. (2017). Social Media spatial monitor of Coup Attempt in the Republic of Turkey. In 2017 17th International Conference on Computational Science and Its Applications (ICCSA), 1–6. IEEE.

Gupta, A., Lamba, H., and Kumaraguru, P. (2013). $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter. *In 2013 APWG eCrime Researchers Summit*. IEEE. doi:10.1109/ecrs.2013.6805772

Haciyakupoglu, G., and Zhang, W. (2015). Social Media and Trust during the Gezi Protests in Turkey. *Journal of Computer-Mediated Communication*, *20*(4), 450–466. doi:10.1111/jcc4.12121

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682–703.

Haklay, M., Basiouka, S., Antoniou, V., and Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, *47*(4), 315–322. doi:10.1179/000870410x12911304958827

Hall, G. B., Chipeniuk, R., Feick, R. D., Leahy, M. G., and Deparday, V. (2010). Community-based production of geographic information using open source software and Web 2.0. *International Journal of Geographical Information Science*, *24*(5), 761–781.

Han, J., Kamber, M., and Tung, A. K. H. (2001). Spatial clustering methods in data mining. *In Geographic Data Mining and Knowledge Discovery*, 188–217. Taylor & Francis. doi:10.4324/9780203468029_chapter_8

Hancilar, U., Tuzun, C., Yenidogan, C., and Erdik, M. (2010). ELER software – a new tool for urban earthquake loss assessment. *Natural Hazards and Earth System Science*, *10*(12), 2677–2696. doi:10.5194/nhess-10-2677-2010

Hara, Y. (2015). Behaviour analysis using tweet data and geo-tag data in a natural disaster. *Transportation Research Procedia.* doi:10.1016/j.trpro.2015.12.033

Hasby, M., and Khodra, M. L. (2013). Optimal path finding based on traffic information extraction from Twitter. *In International Conference on ICT for Smart Society*, 1–5, IEEE.

Hassan, S., Rafi, M., and Shaikh, M. S. (2011). Comparing SVM and naive Bayes classifiers for text categorization with Wikitology as knowledge enrichment. *In 2011 IEEE 14th International Multitopic Conference*. IEEE. doi:10.1109/inmic.2011.6151495

Healy, M., Delany, S., and Zamolotskikh, A. (2005). An assessment of case-based reasoning for short text message classification. *Procs. of 16th Irish Conference on Artificial Intelligence and Cognitive Science*, 257–266.

Hearst, M. A. (1999). Untangling text data mining. *In Proceedings of the 37th Annual meeting of the Association for Computational Linguistics*. doi:10.3115/1034678.1034679

Hecht, B., and Shekhar, S. (2014). From GPS and Google Maps to Spatial Computing. Retrieved May 1, 2014, from https://www.coursera.org/course/spatialcomputing

Hecht, B., and Stephens, M. (2014). A tale of cities: Urban biases in volunteered geographic information. *Proceedings of the 8th International Conference on Weblogs and Social Media*.

Hirata, E., Giannotti, M. A., Larocca, A. P. C., & Quintanilha, J. A. (2015). Flooding and inundation collaborative mapping - use of the

Crowdmap/Ushahidi platform in the city of Sao Paulo, Brazil. *Journal of Flood Risk Management.* doi:10.1111/jfr3.12181

Hochenbaum, J., Vallis, O. S., and Kejariwal, A. (2017). Automatic anomaly detection in the cloud via statistical learning. *ArXiv Preprint* ArXiv:1704.07706.

HOT. (2020). Disaster Response. Retrieved May 10, 2020, from https://www.hotosm.org/impact-areas/disaster-response/

Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein, H. M., Halliwell, M. R., …, and Griffith, S. A. (2015). Social media and disasters: A functional framework for social media use in disaster planning, response, and research. *Disasters*. doi:10.1111/disa.12092

Hu, T., Yang, J., Li, X., and Gong, P. (2016). Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sensing*, *8*(2), 151. doi:10.3390/rs8020151

Hu, Y., John, A., Wang, F., and Kambhampati, S. (2012). ET-LDA: Joint topic modeling for aligning events and their twitter feedback. *Proceedings of the National Conference on Artificial Intelligence*.

HUBÁLEK, Z. (1982). COEFFICIENTS OF ASSOCIATION AND SIMILARITY, BASED ON BINARY (PRESENCE-ABSENCE) DATA: AN EVALUATION. *Biological Reviews*, *57*(4), 669–689. doi:10.1111/j.1469-185x.1982.tb00376.x

Hürriyet. (2016). Dakika dakika darbe girişimi - 15-16 Temmuz 2016. Retrieved from http://www.hurriyet.com.tr/dakika-dakika-darbe-girisimi-15-16-temmuz-2016-40149409

IFRC. (2020). What is a disaster? Retrieved February 1, 2020, from https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster/

Ikonomakis, M., Kotsiantis, S., and Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS Transactions on Computers*. doi:10.11499/sicejl1962.38.456

Ilina, E., Hauff, C., Celik, I., Abel, F., and Houben, G.J. (2012). Social Event Detection on Twitter. *In Lecture Notes in Computer Science*, 169–176. Springer Berlin Heidelberg. doi:10.1007/978-3-642-31753-8_12

Ishino, A., Odawara, S., Nanba, H., and Takezawa, T. (2012). Extracting transportation information and traffic problems from tweets during a disaster. *Proc. IMMM*, 91–96.

Issa, E., Tsou, M. H., Nara, A., and Spitzberg, B. (2017). Understanding the spatio-temporal characteristics of Twitter data with geotagged and non-geotagged content: two case studies with the topic of flu and Ted (movie). *Annals of GIS*, *23*(3), 219–235. doi:10.1080/19475683.2017.1343257

Iwanaga, I. S. M., Nguyen, T. M., Kawamura, T., Nakagawa, H., Tahara, Y., and Ohsuga, A. (2011). Building an earthquake evacuation ontology from twitter. In 2011 IEEE International Conference on Granular Computing, 306–311. doi:10.1109/GRC.2011.6122613

Johnson, R. (2000). GIS technology for disasters and emergency management. An ESRI White Paper.

Jung, M. G., Youn, S. A., Bae, J., and Choi, Y. L. (2015). A Study on Data Input and Output Performance Comparison of MongoDB and PostgreSQL in the Big Data Environment. *In 2015 8th International Conference on Database Theory and Application (DTA)*. IEEE. doi: 10.1109/dta.2015.14

Karaman, H., and Erden, T. (2014). Net earthquake hazard and elements at risk (NEaR) map creation for city of Istanbul via spatial multi-criteria decision analysis. *Natural Hazards*, *73*(2), 685–709.

Karaman, H., Şahin, M., Elnashai, A. S., and Pineda, O. (2008). Loss assessment study for the Zeytinburnu district of Istanbul using Maeviz-Istanbul (HAZTURK). *Journal of Earthquake Engineering*, *12*(S2), 187–198.

Kaya, M., Fidan, G., and Toroslu, I. H. (2012). Sentiment Analysis of Turkish Political News. *In 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. IEEE. doi:10.1109/wi-iat.2012.115

Khorram, Y. (2012). As Sandy pounded NYC, fire department worker was a Twitter lifeline. *CNN*. Retrieved from http://edition.cnn.com/2012/11/01/tech/social-media/twitter-fdny/

Kim, K. S., Kojima, I., and Ogawa, H. (2016). Discovery of local topics by using latent spatio-temporal relationships in geo-social media. *International Journal of Geographical Information Science*, *30*(9), 1899–1922.

Kircher, C. A., Whitman, R., and Holmes, W. T. (2006). HAZUS Earthquake Loss Estimation Methods. *Natural Hazards Review*, *7*(2), 45–59. doi:10.1061/(asce)1527-6988(2006)7:2(45)

KNIME. (2014). KNIME Twitter Nodes. Retrieved from https://www.knime.org/blog/knime-twitter-nodes

Kocaman, S., Anbaroglu, B., Gokceoglu, C., and Altan, O. (2018). A review on citizen science (CitSci) applications for disaster management. *Int Arch Photog Rem Sens Spatial Inf Sci*, *42*(3), W4.

Konukcu, B. E., Karaman, H., and Şahin, M. (2016). Building damage analysis for the updated building dataset of Istanbul. *Natural Hazards*. doi:10.1007/s11069-016-2530-7

Kudou, T. (2013). mecab. Retrieved from https://github.com/taku910/mecab

Kulczyński, S. (1928). Die pflanzenassoziationen der pieninen.

Landwehr, P. M., Wei, W., Kowalchuck, M., and Carley, K. M. (2016). Using tweets to support disaster planning, warning and response. *Safety Science*, *90*, 33–47. doi:10.1016/j.ssci.2016.04.012

Lang, W. S., and Wilkerson, J. (2008). Accuracy vs. Validity, Consistency vs. Reliability, and Fairness vs. Absence of Bias: A Call for Quality. *Online Submission*.

Lansley, G., and Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, *58*, 85–96.

Lawrence, L. (2014). Reliability of sentiment mining tools: a comparison of semantria and social mention (Bachelor's thesis, University of Twente).

Lee, K., Ganti, R., Srivatsa, M., and Mohapatra, P. (2013). Spatio-temporal provenance: Identifying location information from unstructured text. *In 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 499–504.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., and Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*. doi:10.5210/fm.v18i5.4366

Lemon, J., Bolker, B., Oom, S., Klein, E., Rowlingson, B., Wickham, H., …, and Toews, M. (2020). Package 'plotrix.'

Lenormand, M., Tugores, A., Colet, P., and Ramasco, J. J. (2014). Tweets on the road. *PloS One*, *9*(8), e105407.

Li, L., Goodchild, M. F., and Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, *40*(2), 61–77.

Lim, M. (2012). Clicks, Cabs, and Coffee Houses: Social Media and Oppositional Movements in Egypt, 2004-2011. *Journal of Communication*, *62*(2), 231–248. doi:10.1111/j.1460-2466.2012.01628.x

Lin, Y.R., and Margolin, D. (2014). The ripple of fear, sympathy and solidarity during the Boston bombings. *EPJ Data Science*, *3*(1), 1–28.

Liu, B., and Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. *In Mining Text Data* 415–463. Springer US. doi:10.1007/978-1-4614-3223-4_13

Ma, J., Gao, W., and Wong, K.F. (2018). Rumor Detection on Twitter with Tree-structured Recursive Neural Networks. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics. doi:10.18653/v1/p18-1184

Malik, M. M., Lamba, H., Nakos, C., and Pfeffer, J. (2015). Population bias in geotagged tweets. *Ninth International AAAI Conference on Web and Social Media*.

Mason, H., and Wiggins, C. (2010). A taxonomy of data science. *Dataists*. Com, *9*.

Mathew, A. B., and Madhu Kumar, S. D. (2015). Analysis of data management and query handling in social networks using NoSQL databases. *In 2015 International Conference on Advances in Computing, Communications and Informatics*. IEEE. doi:10.1109/icacci.2015.7275708

McClendon, S., and Robinson, A. C. (2012). Leveraging geospatially-oriented social media communications in disaster response. *ISCRAM 2012 Conference Proceedings - 9th International Conference on Information Systems for Crisis Response and Management*.

Meier, P. (2012). Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time. *Journal of Map and Geography Libraries*, *8*(2), 89–100. doi:10.1080/15420353.2012.663739

Meier, P., and Werby, O. (2011). Ushahidi in Haiti: The use of crisis mapping during the 2009 earthquake in Haiti. *Proceedings of the IADIS International Conference e-Democracy, Equity and Social Justice*, 247–250.

Mendoza, M., Poblete, B., and Valderrama, I. (2019). Nowcasting earthquake damages with Twitter. *EPJ Data Science*, *8*(1). doi:10.1140/epjds/s13688-019-0181-0

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2018). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2017. *R Package Version*, *1*(8).

Michael, E. L. (1920). Marine Ecology and the Coefficient of Association: A Plea in Behalf of Quantitative Biology. *The Journal of Ecology*, *8*(1), 54. doi:10.2307/2255213

Middleton, S. E., Kordopatis-Zilos, G., Papadopoulos, S., and Kompatsiaris, Y. (2018). Location Extraction from Social Media. *ACM Transactions on Information Systems*, *36*(4), 1–27. doi:10.1145/3202662

Middleton, S. E., and Krivcovs, V. (2016). Geoparsing and Geosemantics for Social Media: Spatiotemporal Grounding of Content Propagating Rumors to Support Trust and Veracity Analysis during Breaking News. *ACM Trans. Inf. Syst.*, *34*(3), 1–26. doi:10.1145/2842604

Middleton, S. E., Middleton, L., and Modafferi, S. (2014). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*. doi:10.1109/MIS.2013.126

Mocnik, F. B., Mobasheri, A., Griesbaum, L., Eckle, M., Jacobs, C., and Klonner, C. (2018). A grounding-based ontology of data quality measures. *Journal of Spatial Information Science*, *16*. doi:10.5311/josis.2018.16.360

Mocnik, F.-B., Mobasheri, A., and Zipf, A. (2018). Open source data mining infrastructure for exploring and analysing OpenStreetMap. *Open Geospatial Data, Software and Standards*, *3*(1), 7.

Moffitt, J. (2017). Tweet Metadata Timeline. Retrieved from http://support.gnip.com/articles/tweet-timeline.html

Mohammad, S. M., and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*. doi:10.1111/j.1467-8640.2012.00460.x

Mooney, P., Corcoran, P., and Winstanley, A. C. (2010). Towards quality metrics for OpenStreetMap. *In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10*. ACM Press. doi:10.1145/1869790.1869875

Morstatter, F., Pfeffer, J., and Liu, H. (2014). When is it biased?: assessing the representativeness of twitter's streaming API. *Proceedings of the 23rd International Conference on World Wide Web*, 555–556.

Morstatter, F., Pfeffer, J., Liu, H., and Carley, K. M. (2013). Is the sample good enough? comparing data from twitter's streaming api with twitter's

firehose. *Seventh International AAAI Conference on Weblogs and Social Media*.

Muller, J.-P., Tao, Y., Sidiropoulos, P., Gwinner, K., Willner, K., Fanara, L., …, and Bamford, S. (2016). EU-FP7-iMARS: ANALYSIS OF MARS MULTI-RESOLUTION IMAGES USING AUTO-COREGISTRATION, DATA MINING AND CROWD SOURCE TECHNIQUES: PROCESSED RESULTS; A FIRST LOOK. ISPRS - *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B4, 453–458. doi:10.5194/isprs-archives-xli-b4-453-2016

Muralidharan, S., Rasmussen, L., Patterson, D., and Shin, J.-H. (2011). Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts. *Public Relations Review*, *37*(2), 175–177. doi:10.1016/j.pubrev.2011.01.010

Murzintcev, N., and Cheng, C. (2017). Disaster Hashtags in Social Media. *ISPRS International Journal of Geo-Information*, *6*(7), 204. doi:10.3390/ijgi6070204

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings*.

NodeXL for Programmers. (2011). In Analyzing Social Media Networks with NodeXL 273–275. Elsevier. doi:10.1016/b978-0-12-382229-1.00025-4

NTV. (2016). 15 Temmuz gecesi ve sonrasında neler yaşandı. Retrieved from https://www.ntv.com.tr/galeri/turkiye/15-temmuz-gecesi-ve-sonrasinda-neler-yasandi,fBIiAr1pu0WkrhUBXdZykA/ClgQwqBRtEaLnVzxTJdIXw

Ord, J. K., and Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*. doi:10.1111/j.1538-4632.1995.tb00912.x

OSM. (2016). Open Street Map About. Retrieved from https://www.openstreetmap.org/about

Ozdikis, O., Oğuztüzün, H., and Karagoz, P. (2017). A survey on location estimation techniques for events detected in Twitter. *Knowledge and Information Systems*. doi:10.1007/s10115-016-1007-z

Ozturk, N., and Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, *35*(1), 136–147. doi:10.1016/j.tele.2017.10.006

Parker, C. J. (2014). A Framework of Neogeography. *In The Fundamentals of Human Factors Design for Volunteered Geographic Information* 11–22. Springer.

Parsons, T. (2004). Recalculated probability of M≥ 7 earthquakes beneath the Sea of Marmara, Turkey. *Journal of Geophysical Research: Solid Earth*, *109*(B5).

Parsons, T., Toda, S., Stein, R. S., Barka, A., and Dieterich, J. H. (2000). Heightened odds of large earthquakes near Istanbul: An interaction-

based probability calculation. *Science*. doi:10.1126/science.288.5466.661

Poell, T., and Borra, E. (2011). Twitter, YouTube, and Flickr as platforms of alternative journalism: The social media account of the 2010 Toronto G20 protests. *Journalism: Theory, Practice & Criticism*, *13*(6), 695–713. doi:10.1177/1464884911431533

Poorazizi, M. E., Hunter, A. J. S., and Steiniger, S. (2015). A Volunteered Geographic Information Framework to Enable Bottom-Up Disaster Management Platforms. *ISPRS International Journal of Geo-Information*, *4*(3), 1389–1422.

PostGIS. (2017). About PostGIS. Retrieved from https://postgis.net/

PostgreSQL. (2017). Homepage. Retrieved from https://www.postgresql.org/

Power, R., Robinson, B., and Wise, C. (2015). Using Crowd Sourced Content to Help Manage Emergency Events. *In Social Media for Government Services* 247–270. Springer.

R Core Team. (2013). R: A language and environment for statistical computing.

Real, R., and Vargas, J. M. (1996). The Probabilistic Basis of Jaccard's Index of Similarity. *Systematic Biology*, *45*(3), 380–385. doi:10.1093/sysbio/45.3.380

Red Hen Systems, Inc. (2016). MapCalcTM Learner-Academic software. Retrieved from http://www.innovativegis.com/basis/Senarios/Materials/MC_instructor/MC_instructor.htm

Ripley, B., Venables, W., and Ripley, M. B. (2016). Package 'nnet.' *R Package Version*, *7*, 3–12.

Rosser, J. F., Leibovici, D. G., and Jackson, M. J. (2017). Rapid flood inundation mapping using social media, remote sensing and topographic data. *Natural Hazards*, *87*(1), 103–120. doi:10.1007/s11069-017-2755-0

Ryerson University. (2016). Social Media Data Stewardship Social Media Research Toolkit. Retrieved from https://socialmediadata.org/social-media-research-toolkit/

Ryoo, K., and Moon, S. (2014). Inferring Twitter user locations with 10 km accuracy. *In Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*. ACM Press. doi:10.1145/2567948.2579236

Sadilek, A., Kautz, H., and Bigham, J. P. (2012). Finding your friends and following them to where you are. *In Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. ACM Press. doi:10.1145/2124295.2124380

Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. *Proceedings of the 19th International Conference on World Wide Web, WWW '10*. doi:10.1145/1772690.1772777

Schneider, P. J., and Schauer, B. A. (2006). HAZUS - Its Development and Its Future. *Natural Hazards Review*, *7*(2), 40–44. doi:10.1061/(asce)1527-6988(2006)7:2(40)

Schroeder, P. (1996). Criteria for the Design of a GIS/2. Specialists' meeting for NCGIA Initiative 19. GIS and Society, Summer.

Scistarters. (2017). Project Finder. Retrieved from https://scistarter.com/finder

Scott, L. M., and Janikas, M. v. (2009). Spatial Statistics in ArcGIS. *In Handbook of Applied Spatial Analysis* 27–41. Springer Berlin Heidelberg. doi:10.1007/978-3-642-03647-7_2

Seker, S. E. (2014). Metin Madenciliği (Text Mining). Retrieved from http://bilgisayarkavramlari.sadievrenseker.com/2014/06/15/metin-madenciligi-text-mining/

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., and Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *In International Journal of Geographical Information Science.* doi:10.1080/13658816.2016.1189556

Sieber, R. (2006). Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers*, *96*(3), 491–507.

Signorini, A., Segre, A. M., and Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS One*, *6*(5), e19467.

Sloan, L., and Morgan, J. (2015). Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PloS One*, *10*(11), e0142209.

Slowikowski, K. (2019). ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2." *R Package Version 0.8.1*.

Smith, J. R., and Chang, S. F. (1996). Automated binary texture feature sets for image retrieval. *In 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. IEEE. doi:10.1109/icassp.1996.545867

Sørensen, T., Sørensen, T. A., Sørensen, T. J., SORENSEN, T., Sorensen, T., Sorensen, T. A., and Biering-Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons.

Sorokin, A., and Forsyth, D. (2008). Utility data annotation with Amazon Mechanical Turk. *In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE. doi:10.1109/cvprw.2008.4562953

Sridhar, R., and Vivek, K. (2015). Unsupervised Topic Modeling for Short Texts Using Distributed Representations of Words. *In Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*. Association for Computational Linguistics. doi:10.3115/v1/w15-1526

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., and Demirbas, M. (2010). Short text classification in twitter to improve information filtering. *In Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. ACM Press. doi:10.1145/1835449.1835643

StanfordNLPGroup. (2008). Stemming and Lemmatization.

Statista. (2017a). Active social network penetration in selected countries as of January 2017. *The Statistics Portal*. Retrieved from https://www.statista.com/statistics/282846/regular-social-networking-usage-penetration-worldwide-by-country/

Statista. (2017b). Forecast of social network user numbers in Turkey from 2015 to 2022 (in million users). The Statistics Portal. Retrieved from https://www.statista.com/statistics/569090/predicted-number-of-social-network-users-in-turkey/

Statista. (2017c). Most famous social network sites worldwide as of September 2017, ranked by number of active users (in millions). The Statistics Portal. Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Statista. (2018). Most popular social networks worldwide as of April 2018, ranked by number of active users (in millions). Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Statista. (2019). Most popular social networks worldwide as of October 2019, ranked by number of active users. Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Statista. (2020). Global digital population as of April 2020 (in millions). Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/

Stephens, M. (2013). Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal*, *78*(6), 981–996. doi:10.1007/s10708-013-9492-z

Sun, Y., Wonk, A. K. C., and Kamel, M. S. (2009). CLASSIFICATION OF IMBALANCED DATA: A REVIEW. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(04), 687–719. doi:10.1142/s0218001409007326

Tamura, K., and Ichimura, T. (2013). Density-Based Spatiotemporal Clustering Algorithm for Extracting Bursty Areas from Georeferenced Documents. *In 2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE. doi:10.1109/smc.2013.356

Tarhan, C., Coşkun, Z., and Zülfikar, C. (2013). DEPREM BİLGİ SİSTEMİ.

Terpstra, T., de Vries, A., and Stronkman, R. (2012). Towards a realtime Twitter analysis during (flood) crises for operational (flood) crisis management. *In Comprehensive Flood Risk Management*. doi:10.1201/b13715-221

Trivedi, M., Sharma, S., Soni, N., and Nair, S. (2015). Comparison of text classification algorithms. *International Journal of Engineering Research & Technology*, *4*(02).

Tsou, M.-H., Zhang, H., and Jung, C.-T. (2017). Identifying data noises, user biases, and system errors in geo-tagged twitter messages (tweets). ArXiv Preprint ArXiv:1712.02433.

Tufekci, Z., and Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication*, *62*(2), 363–379. doi:10.1111/j.1460-2466.2012.01629.x

Tulloch, D. L. (2008). Is VGI participation? From vernal pools to video games. *GeoJournal*, *72*(3–4), 161–171.

Turkish Statistical Institute. (2019). Main Statistics, Population and Demography, Population Statistics, Population of Provinces by Years. Retrieved from http://www.tuik.gov.tr/UstMenu.do?metod=temelist

Turner, A. (2006). Introduction to neogeography. O'Reilly.

Twitter. (2015). AnomalyDetection R package.

Twitter. (2020a). API Reference Index. Retrieved from https://developer.twitter.com/en/docs/api-reference-index

Twitter. (2020b). Products for researchers. Retrieved from https://developer.twitter.com/en/use-cases/academic-researchers/products-for-researchers

USGS. (2019). DYFI Summary Maps.

Ushahidi. (2017). Read The Crowd. Retrieved from https://www.ushahidi.com/

Utani, A., Mizumoto, T., and Okumura, T. (2011). How geeks responded to a catastrophic disaster of a high-tech country. *In Proceedings of the Special Workshop on Internet and Disasters - SWID '11. ACM Press*. doi:10.1145/2079360.2079369

Valenzuela, S., Arriagada, A., and Scherman, A. (2012). The Social Media Basis of Youth Protest Behavior: The Case of Chile. *Journal of Communication*, *62*(2), 299–314. doi:10.1111/j.1460-2466.2012.01635.x

Vallis, O. S., Hochenbaum, J., and Kejariwal, A. (2014). AnomalyDetection: Anomaly Detection Using Seasonal Hybrid Extreme Studentized Deviate Test. *R Package Version*, *1*.

Vo, D. T., and Zhang, Y. (2015). Target-dependent twitter sentiment classification with rich automatic features. *IJCAI International Joint Conference on Artificial Intelligence*.

Vural, A. G., Cambazoglu, B. B., Senkul, P., and Tokgoz, Z. O. (2012). A Framework for Sentiment Analysis in Turkish: Application to Polarity Detection of Movie Reviews in Turkish. *In Computer and Information Sciences III* 437–445. Springer London. doi:10.1007/978-1-4471-4594-3_45

Wald, D. J., Quitoriano, V., Worden, C. B., Hopper, M., and Dewey, J. W. (2012). USGS "Did You Feel It?" Internet-based macroseismic intensity maps. 2012, 54(6). doi:10.4401/ag-5354

Wang, B., and Zhuang, J. (2018). Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. *Natural Hazards*, *93*(3), 1145–1162.

Wang, Z., Ye, X., and Tsou, M. H. (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Natural Hazards*, *83*(1), 523–540. doi:10.1007/s11069-016-2329-6

Wardlaw, J, and Jackson, B. (2018). RIGHTS LAB: SLAVERY FROM SPACE. Retrieved from https://www.zooniverse.org/projects/ezzjcw/rights-lab-slavery-from-space

Wardlaw, J., Sprinks, J., Houghton, R. J., Bamford, S., Muller, J.-P., Gasselt, S. …, and Kim, J. (2018). MARS IN MOTION. Retrieved from https://www.zooniverse.org/projects/imarsnottingham/mars-in-motion

WHO, and EHA. (1999). EMERGENCY HEALTH TRAINING PROGRAMME FOR AFRICA. Retrieved from http://apps.who.int/disasters/repo/5512.pdf

Wickham, H., and Grolemund, G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. *In Southeast Asian Journal of Tropical Medicine and Public Health*.

Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., …, and Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*. doi:10.1007/s10115-007-0114-2

Yamamato, Y. (2015). Twitter 4j.

Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R. (2012). Using social media to enhance emergency situation awareness. IEEE *Intelligent Systems*, *27*(6), 52–59. doi:10.1109/MIS.2012.6

Zhao, S., Zhong, L., Wickramasuriya, J., and Vasudevan, V. (2011). Human as Real-Time Sensors of Social and Physical Events: A Case Study of Twitter and Sports Games. *CoRR*, abs/1106.4.

Zi, C., Steven, G., Haining, W., and Sushil, J. (2010). Who is tweeting on Twitter: human, bot, or cyborg? *In Proceedings of the 26th Annual Computer Security Applications Conference 978-1-4503-0133-6* 21–30. ACM. doi:10.1145/1920261.1920265

Zooniverse. (2017). People - powered research. Retrieved from https://www.zooniverse.org/

Zou, L., Lam, N. S. N., Shams, S., Cai, H., Meyer, M. A., Yang, S., …, and Reams, M. A. (2019). Social and geographical disparities in Twitter use during Hurricane Harvey. *International Journal of Digital Earth*, *12*(11), 1300–1318. doi:10.1080/17538947.2018.1545878

CURRICULUM VITAE

Name Surname **:** Ayse Giz Gulnerman

Place and Date of Birth **:** ADANA – 26.08.1986

E-Mail **:** ggulnerman@gmail.com

EDUCATION **:**

- **B.Sc.** **:** 2009, Middle East Technical University, Faculty of Architecture, City and Regional Planning Department
- **B.Sc. (Erasmus)** **:** 2008 (January - July), The University of Manchester, Planning and Landscape Department, Manchester- UK
- **M.Sc.** **:** 2014, Istanbul Technical University, Civil Engineering Faculty, Geomatics Engineering Department
- **Ph.D. (Visiting Scholar):** 2017 - 2018, University of Nottingham, Nottingham Geospatial Institute (NGI), Nottingham - UK

PROFESSIONAL EXPERIENCE:

- 2010 - 2013 Private Sector Experience on Geographic Information Systems
- 2013 - 2014 Research Assistant at Gazi University
- 2014 - ____ Research Assistant at Istanbul Technical University

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Gulnerman, A.G.,** Karaman, H. 2015: PPGIS Case Studies Comparison and Future Questioning. The 15th International Conference on Computational Science and Its Applications (ICCSA 2015), June 22-25, 2015 Banff- Calgary- Canada. doi: 10.1109/ICCSA.2015.15
- **Gulnerman, A.G.,** Gengec, N. E., Karaman, H. 2016: Review of Public Tweets Over Turkey Within a Pre-Determined Time. the 1st International Conference on Smart Data and Smart Cities, September 7-9, 2016 Split- Croatia. doi:10.5194/isprs-annals-III-4-W1-153-2016.
- **Gulnerman, A.G.,** Karaman, H. 2017: Social Media spatial monitor of Coup Attempt in the Republic of Turkey. The 17th International Conference on Computational Science and Its Applications (ICCSA 2017), July 3-7, 2017 Trieste- Italy. doi: 10.1109/ICCSA.2017.7999644

- **Gulnerman, A.G.,** Basiri, A., Karaman, H., Marsh, S. 2018: Geography of pre-referendum tweets and Twitter users, GISRUK 2018- 26th GIScience Research UK Conference, University of Leicester- UK
- **Gulnerman, A.G.,** Karaman, H., Basiri, A., Marsh, S. 2018: Credibility Index of the Social Media Users to Detect Spatial Outliers, The RGS-IBG 2018 Annual International Conference, Cardiff University, August 28-31 ,2018 Cardiff, Wales-UK
- **Gulnerman, A.G.,** Karaman, H., 2020: Spatial Reliability Assessment of Social Media Mining Techniques with regard to Disaster Domain-based Filtering, *ISPRS International Journal of Geo-Information*, 9, 245. doi: https://doi.org/10.3390/ijgi9040245
- **Gulnerman, A.G.,** Karaman, H., Pekaslan, D., Bilgi, S., 2020: Citizens' Spatial Footprint on Twitter: Anomaly, Trend and Bias Investigation in Istanbul, *ISPRS International Journal of Geo-Information*, 9, 222. doi: https://doi.org/10.3390/ijgi9040222
- **Gulnerman, A. G.**, Karaman, H., Basiri, A. 2020: New age of crisis management with social media, *Open Source Geospatial Science for Urban Studies* - Lecture Notes in Intelligent Transportation and Infrastructure, Springer, ISBN: 978-3-030-58232-6. doi: https://doi.org/10.1007/978-3-030-58232-6_8
- **Gulnerman, A.G.,** Karaman, H. 2020: Sosyal Medyanın Gönüllü Coğrafi Veri Olarak Kullanımı ve Sosyal Medya Verilerinden Coğrafya Sözlüğü Üretimi. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 20 (2), 276-286. Retrieved from https://dergipark.org.tr/tr/pub/akufemubid/issue/54356/667397

OTHER PUBLICATIONS AND PRESENTATIONS:

- **Gulnerman, A.G.,** Bük, O., Göksel, Ç., 2013: Afet Sonrası Kriz Yönetiminde Hayat Kurtaran Büfe Önerisi. TMMOB Coğrafi Bilgi Sistemleri Kongresi, Kasım 11-13, 2013 Ankara, Türkiye.
- **Gulnerman, A.G.,** Tezer, A., Göksel, Ç., 2014: Kriz Yönetiminde Konumsal Tabanlı Bir İletişim Modeli Önerisi Ve Mevcut Uygulamalarla Karşılaştırılması. 5. UZAKTAN ALGILAMA-CBS SEMPOZYUMU, Ekim 14-17, 2014 İstanbul, Türkiye.
- **Gulnerman, A.G.,** Goksel, C., 2015: Life Saving Kiosk for Sustainable Disaster Crisis Management. the 18th International Symposium on Environmental Pollution and its Impact on Life in the Mediterranean Region (MESAEP 2015), September 26-30, 2015 Heraklion- Crete- Greece.
- **Gulnerman, A.G.,** Gokkaya, M., Demirel, H., 2015: Geovisualization of Time Variance. the 18th International Symposium on Environmental Pollution and its Impact on Life in the Mediterranean Region (MESAEP 2015), September 26-30, 2015 Heraklion- Crete- Greece.
- Ozturk, O., Bilgi, S., **Gulnerman, A.G.,** 2016: Analysing Spatial Distribution of OSYM Offices in Istanbul due to the Service Areas, International Scientific Conference on Applied Sciences 2016, September 27-30 2016- Turkey.

- Ozturk, O., **Gulnerman, A.G.,** Kalkan Y., Bilgi, S., 2016: Analysing Impact Area of OSYM Offices in Istanbul by IDW Method, American Geophysical Union: AGU 2016, December 12-16 2016- USA.
- Bilgi, S., Ozturk, O., **Gulnerman, A.G.**, 2017: Navigation System for Blind, Hearing and Visually Impaired People İn ITU Ayazaga Campus, 2017 International Conference on Computing Networking and Informatics (ICCNI), October 2017, Covenant University, Ota, Nigeria, doi: 10.1109/ICCNI.2017.8123814
- **Gulnerman, A.G.,** Yavasoglu H., Alkan, M.N., Ozsoy, B., Durak O.S., Oktar, O., 2017: The Use and Importance of Geographic Information Systems in Polar Studies, Polar Science Program Workshop, Apr 12-13, 2017 Istanbul, Turkey.
- Yavasoglu H., Alkan, M.N., **Gulnerman, A.G.,** Ozsoy, B., Basar E., 2017: Geodetic Measurements in Antarctica in Different Geographical Regions, Polar Science Program Workshop, Apr 12-13, 2017 Istanbul, Turkey.
- Alkan, M.N., Yavasoglu H., **Gulnerman, A.G.,** Ozsoy, B., 2017: GNSS Studies in Antarctica and Horseshoe Case, Polar Science Program Workshop, Apr 12-13, 2017 Istanbul, Turkey.
- Karacik B., **Gulnerman, A.G.,** Yılmaz A., Yakan S.D., Tutak B., Okay O., 2017: Propagation of Persistent Organic Pollutants in Poles, Polar Science Program Workshop, Apr 12-13, 2017 Istanbul, Turkey.
- **Gulnerman, A.G.,** Goksel, C., Tezer, A., 2017: Disaster Capacity Building with a GIS Tool of Public Participation. Fresenius Environmental Bulletin, 26(1), 237-243.
- Bilgi, S., **Gulnerman, A.G.,** Arslanoglu, B., Karaman, H., Ozturk, O., 2019: Complexity Measures of Sport Amenities Allocation in Urban Area by Metric Entropy and Public Demand Compatibility. International Journal of Engineering and Geosciences, 4(3), 141-148, doi: https://doi.org/10.26833/ijeg.540180.
- Yavasoglu, H.H., Karaman, H., Ozsoy, B., Bilgi, S., Tutak, B., **Gengec, A.G.G.,** Oktar, O., Yirmibesoglu, S., 2019: Site selection of the Turkish Antarctic Research station using Analytic Hierarchy Process. Polar Science, 22, doi: https://doi.org/10.1016/j.polar.2019.07.003.