UNIVERSITY OF LEEDS

On The Correlates of Group-based Emotions of Social Movements in Social Media.

Daniel Valdenegro Ibarra

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds

Faculty of Social Sciences

School of Politics and International Studies

November 2022

Intellectual Property

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Daniel Hernán Valdenegro Ibarra to be identified as Author of this work has been asserted by his in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

First I would like to thank my esteemed supervisors, Prof. Viktoria Spaiser, Prof. Richard P. Mann and Prof. Jocelyn Evans for their invaluable advice and continuous support and patience during my PhD study. If this project had any hope to come to a fruitful conclusion it was largely thanks to their help and guidance. I would also like to thank Prof. Serge Sharoff for his technical advice for which without this thesis would have taken a very different path. Additionally, this endeavour would have not been possible without the generous support of the Advance Human Capital programme of the Chilean National Research and Development Agency (ANID), who financed my research.

I would like to thank Dr. Nicole Nisbett for her feedback and moral support, but more importantly, for being my friend and the best office mate anyone could hope for.

I would like to thank my parents and sister, who were always there to keep my spirits and motivation high during this process. I would also like to thank my cats, Mushu and Poki, for all their emotional support.

Finally, I would like to thank my spouse and best friend Tania for always believing in me. Without her love, patience and support I would have never been able to complete this project.

Abstract

This project explores the application of large language models for detecting emotions and predicting the collective actions within the context of three distinct social movements: Fridays For Future, Hong Kong 2019 Social Movement and the Chilean 2019 Social Movement. Using large amounts of text data from Twitter, I examine the relationship between emotions expressed in tweets and the occurrence of violent and non-violent collective actions. The study focuses on four key emotions: anger, fear, joy, and sadness, which I argue are foundational in human perception and motivation. This project attempts to bridge the gap between the large quantities of social media data related to social movements currently available and previous socio-psychological theories of participation in collective actions.

My analysis suggests that anger is the main expressed emotion in the data of all three social movements. However, there is significant variation on what is the most significant second emotion across social movements. Furthermore, several emotions were found to be predictive of either violent or non-violent collective actions, although these relationships are not consistent across social movements. Finally, analysis shows a strong feedback loop, computed using auto-correlation analysis, between reports of collective actions and the occurrence of future collective actions.

These findings contribute to the understanding of how emotions manifest and influence the social movement's collective actions, and how the reporting of collective actions itself could influence the occurrence of future collective actions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Social movements are a very interesting and important topic for social scientists. They are a collective behaviour of individuals aimed to achieve societal changes. A motor of social adjustment from within the society itself that can lead to great improvements (e.g., environmentalist movements, civil rights movements). These societal improvements, however, usually come after significant efforts from the participants, and sometimes, at very high personal costs. Understanding better what motivates participants of social movements to engage in such costly actions, and predict when those actions are more likely to occur, has been the topic of a great deal of interdisciplinary research over the years, with various degrees of success. In this thesis I will attempt that yet again, this time using the digital traces of the emotions of social movement on social media (specifically Twitter) and connect that data with the reports of their collective actions on news websites.

Twitter data were collected from the Twitter streaming API and it was organised in time-series format, allowing for the description of the evolution over time of the emotions of each social movement.

This work crosses disciplines between machine learning, natural language processing, political science and political psychology, and it is meant to be a cross-disciplinary effort to apply, with the perspective of a social scientist, machine learning and natural language processing techniques to a complex social phenomenon, and carefully evaluate the advantages and disadvantages of said techniques.

A review of the relevant literature and ideas of *emotion* and how they are related to social move-

ments can be found in Chapter 2. In this chapter I review what is understood by emotions, its neurological correlates, its influence in the motivational process and how the concept of emotion has been used in the political psychology literature to model the motivation to participate in activism and collective actions. I close the chapter proposing a framework to investigate four core functional aspects of any social movement: a "Group-based Motivator" that motivates the emergence of the movement; the "Collective Actions" of the movement; what the movement considers its "Ingroup"; and what the movement considers its "Outgroup". Emotions are not expressed in the vacuum. They are directed towards something (e.g., a situation, a person, a group of people, etc.). I argue that participants of social movements should, at least, have some awareness about these four basic aspects, and hence be able to feel emotions towards them. Given this, correctly identifying these aspects in every social movement is crucial to create the profile of emotions of its participants, in relation to the social movement they are part of. Chapter 3 provides a description of the social movements studied.

Chapter 4 provides, first, a detailed description of the Twitter data collection and data cleaning process, as well as a description of the data used in the training process of the machine learning methods used. The later subsection dives into a description of the machine learning and natural language processing techniques used in this project, with a focus on their mathematical and algorithmic foundations, but adapted to be presented and understood to social science audiences.

Finally, Chapter 5 describes the results of the models selection study, the data augmentation study, the emotional classification study over the main Twitter data and the time-series analysis linking the identified emotions with the real-life events of each social movement. The model selection study describes the process of testing the ability of the techniques presented in Chapter 4 to accurately classify the emotions in the tweets posts. The data augmentation study describes the process of using synthetic data in order to improve the accuracy of the techniques used for emotion classification. The emotional classification study describes the results obtained from the application of the emotion detection model to the main social movement Twitter data, by presenting the evolution of the emotional patterns for each of the social movements studied. Finally, the detected emotions are linked with news data (see Section 4.1.2) of real-life events of each social movement, using Vector Auto-Regressive models described in Chapter 4.

1.1 Motivation

The experience of being a participant or a bystander of collective actions has an affective component. Emotions can act as antecedents or consequences of the participation and it is expected that a variety of emotions can be experienced at the same time while being involved in these activities. These emotions provide the participants with the motivation to take part in collective actions or social movements. However, contrary to the opinion of several theoretical models (Dalgleish 2004), emotions do not render participants irrational, and excluding emotions from the models of collective action/social movements will likely limit our understanding of them.

Emotions are a key component of the neuro-psychological motivational process (Dalgleish 2004) as well as being one of the first reactions in response to any new stimuli (Dalgleish 2004; Marcus et al. 2022). In addition, it is well known that information processing through the limbic system —which controls the emotional response— is much faster and occurs before any rational information processing (Marcus et al. 2022). This has led to the hypothesis that emotional information processing can influence decision making, while unconscious rationalisation processes hide this from the subject (Arzheimer et al. 2017; Marcus et al. 2022). This is especially true when individuals are confronted with situations evaluated as threats to their well-being, like social inequalities, unemployment, discrimination, migration, or others (Arzheimer et al. 2017; Marcus et al. 2022), making the study of emotions in the context of collective actions/social movements very relevant to understand the dynamics of participation and the emergence of extreme, radical or violent behaviour within social movements.

Accordingly, the majority of the current psychological models trying to predict participation in collective actions have an emotional component. The Elaborated Social Identity model of Drury and Reicher (1999) and the Social Identity Model of Collective Actions of Van Zomeren, Postmes, et al. (2008) both include the perception of group injustices as main predictor, which leads to emotional reactions (anger, sadness, etc.), motivating the engagement of the participants in collective actions. This was later expanded and refined in the work of Van Zomeren, Leach, et al. (2012) with the Dynamic Dual Pathway Model, in which group-based anger is explicitly considered as one of the most proximal predictors of collective action. However, these models only describe how one emotion, namely anger, is related to the participation in collective actions and activism, but nothing is said about what other emotions might be involved or

expressed by social movements' participants.

Additionally to the above-mentioned limitations, important methodological restrictions can be identified in this research domain. A relevant issue in the case of collective action is the distance between the measurement of emotions and the actual phenomenon, collective actions. The use of self-reported questionnaires and measures of intention of participation are preferred over measures of observed behaviour, mainly because of its straightforward application and practicality. While this is acceptable due to the difficulties in collecting data of actual participation, it cannot be denied that this might bias claims that these studies make over social movements and collective actions. Self-reported measures suffer from social desirability bias, unreliable reporting and self-selection bias. These biases usually affect the expression of emotions and the assessment of the participants. Social desirability will tend to inhibit the expression of emotions deemed socially unacceptable, while self-selection bias in the survey participation could lead to over-represent the opinions and emotions of a very specific group of individuals who do not necessarily represent the whole of the social movement. All of this affects the quality of the data we can obtain from social movements using more traditional methods.

Interestingly, the arrival of inter-connectivity and social media can open opportunities to deal with the methodological restrictions of the collective actions research field. For instance, the digital activities of participants on social media used by social movements leave traceable records. This provides a register of the social movement's participants' behaviour on the internet, in the form of comments, blogs, or micro-activities (e.g., tweets). What makes this type of data relevant for the purposes of this project is its closeness to the individual. This is first hand data of the individual activities and behaviours. Descriptions about their attitudes and emotional states made by themselves without the filter —and the possible bias— of interviewers and without the restrictions of questionnaires. The veil of anonymity provided on many social media platforms also reduces the pressure of social desirability. This has limitations, of course. The main one is that not everyone is on social media, and thus, it is not possible to claim that the data are in any way representative of the target population nor the target social movements under study. Additionally, not all the people that are on social media want to share their emotions related to their participation in social movements, limiting the scope of the conclusions we can draw from this data. However, in this case, and for the purpose of this project, the advantages of social media data outweighs the limitations, as it provides a rich record of online

activities of potentially millions of users associated with social movements overtime. This is something that is extremely difficult to collect from traditional methods. Moreover, the very same architecture of the social media provides us with metadata to assess the popularity, spread and rate of adoption of topics, opinions or even emotions, if we can identify them.

Consequently, since there is not yet a clear understanding of what emotions other than anger are expressed by members of social movements and how these emotions might be related to the real life events of each social movements (e.g., demonstration, strikes, sit-ins, etc.) in this thesis:

- I will investigate the dynamic expression of four emotions (anger, fear, sadness and joy) in the tweets associated with three social movements: Chilean 2019 Social Movement, Hong Kong 2019 Social Movement and Fridays for Future 2019.
- Additionally, I will analyse the relation of the expression of these four emotions with the real-life events and collective actions performed by each of these social movements.

The theoretical reasons for including only anger, fear, sadness and joy as the emotions to be measured in this thesis are described in detail in Section 2.3.1 of Chapter 2. In general terms, previous work in the field of neuro-biology and neuro-psychology suggests that these four emotions have a very strong physiological component (Ekman 1992; Damasio et al. 2000; Dalgleish 2004), making them very foundational. Foundational emotions have a higher chance to be reflected in the subjects' behavioural expressions (e.g., physical expression, vocal expression, textual expression etc.) and with greater clarity, making them ideal for situations where they are measured indirectly (Ekman 1992; Jack, Garrod, and Schyns 2014), as is the case in this research project. So, these emotions are being selected because foundational emotions are likely to be easier to detect in short textual social media data.

1.2 Expected Contributions

This research project is cross-disciplinary. It attempts to combine the theoretical understanding of the social sciences, with most recent advances in Natural Language Processing to try to describe complex social phenomena. This approach can prove to be valuable today when there has been an increasing interest from researchers outside the realm of traditional social sciences to study social problems using mathematical and computation methods, but without a full

understanding of the contexts, antecedents and possible theoretical implications of their findings. This has led to a corpus of literature full of innovative methodologies but with interpretations of the findings that ignore the research made in the same field by social sciences. On the other hand, within the social sciences there is still reticence towards the use of computational methods and the advantages that these can provide, favouring more traditional methodologies, like interviews, surveys, and ethnography (Edelmann et al. 2020). While these methods provide extreme detail and granular information about social phenomena, they require a significant amount of effort and resources from the researchers, yielding only very specific results that are difficult to generalise. Hopefully, this research will help to minimise this reticence, providing results with theoretically-grounded interpretations for an important social phenomena but using replicable computational procedures and techniques. Finally, and most importantly, this project aims to provide a better understanding of how social movements interact and feed from the interactions in social media. This thesis is, in essence, a description of the process of developing and applying reliable techniques for emotions detection in short text (e.g., tweets); a description of the process of applying said techniques over a large corpus of tweets related to several distinct social movements; And finally, a description of how the application of this natural language processing techniques could provide new avenues for the study of social movements and/or other social events using digital footprint data and unstructured data sources like text.

Chapter 2

Theoretical Framework

In this chapter I will present the conceptual antecedents that have led me to propose this research. I will first review some of the major sociological models on social movements and collective action and how they are challenged by the advent of social movements organised using social media, and then proceed to review some key concepts. Finally, I will review some of the most recent socio-psychological models that try to explain participation in collective action from an individual perspective, focusing on the construction of emotions in these models.

2.1 The role of emotions in today's social movements

Sidney Tarrow, in his book *Power in Movement* offer a very good summary of the most relevant sociological traditions in the study of social movements and collective actions of the twentieth century (Tarrow 1998), including the Grievances and Collective Behaviour Theory, the Rational Choice and Resource Mobilisation Theory (Jenkins 1983), Framing and Collective identity Theory (Bartholomew and Mayer 1992), and Political Process Theory (Goodwin and Jasper 1999). I will review briefly the contributions of each theory to the general understanding of social movement and collective actions. For Tarrow, Grievances and Collective Behaviour Theory, which has its period of highest popularity around the 1950s and early 1960s, was the result of the influence of social psychology over American sociologists, who took the concept of "collective behaviour" and applied it to social movements. This led to the conceptualisation of social movements as "emergent phenomena", meaning that they emerge from the behaviour of the individual agents, which are not necessarily conscious of the coordination they are achieving

(e.g., like schooling in fish or flocking in birds). As a consequence of this, social movements were seen as something abnormal, a symptom of a dysfunctional society or the consequences of individual deprivation (Gurr 2015). Tarrow proposed that the vision of social movements as an abnormality, and the lack of an explicit relationship between them and politics, was one the main reason why this theory lost popularity during the 1970s, a decade where social mobilisation was tightly related to political issues (see Tarrow (1998), pp.22–23).

Rational Choice and Resource Mobilisation Theory was born during the late 1960s and early 1970s, during the boom of the anti-war in the United States, powered by young middle class people. This was also the time in which economics was starting to be considered the "best" social science, boosted by the popularity and assumed superiority of its methods and the growing impact of finances over the social lives of individuals (Coase 1977; Freeman 1999; Fourcade et al. 2015). In contribution to this conversation, the economist Mancur Olson published his book *Logic of Collective Action* (Olson 1971) in which he presented a model which, in the opinion of Tarrow, put the problem of collective action on the same level as the problem of marketing: how to attract the biggest amount of participants to achieve a collective good (see Tarrow (1998), p.23). Olson proposed that in large groups, the tendency to "free ride", this being the tendency to drop off from participation in the social movements, was greater due to rational evaluation that one's efforts can be easily replaced by some other's. This tendency could be controlled by offering selective incentives to participants or by imposing participation rules on the members (see Olson (1971), pp. 60-61). However, this "rational choice" model was failing to explain the contemporary social movements, in which thousands of people were protesting and demonstrating on behalf of the interests of other people (see Tarrow (1998), p.23). John McCarthy and Mayer Zald proposed an explanation to this apparent selfless participation. For McCarthy and Zald, the surplus of resources in industrialised societies led to the professionalisation of social movements, promoting the creation of organisations which administered the resources of social movements and defined their goals and methods. These organisations would be ones in charge of bringing together logistic and economic resources and to "call-in" participants when the occasion required mass collective actions. Current examples of such social movement organisations are Amnesty International, Black Lives Matter and Extinction Rebellion. (McCarthy and Zald 1977).

Framing and Collective identity Theory emerged as a response to the emphasis on organisa-

tions and lack of inclusion of individual's emotions and grievances in the Resource Mobilisation Theory. This perspective proposes that social movements were carriers of meaning for the participants, reflecting some of their identity. In this perspective participants can "construct" their participation in the social movement as well as "construct" their subjectivity (see Tarrow (1998), p.26; Benford and Snow (2000)). This perspective draws from psychology and social psychology, in which the construction of identity and self, and the relation between the individual identity and a collective identity, are important factors in the models predicting participation in collective actions. Social Identity Theory proposes that what we know as "identity" is the results of the actions and accommodations we take in order to better fit in a particular social group. These accommodations lead to the construction of self-concepts around a membership in a social group (e.g., "I'm a member of X association," "As a member of X I must be faithful to Y set of values."). A more detailed explanation about Self, Identity and Social Identity Theory can be found in sections 2.3.4 and 2.3.3 of this Chapter.

Finally, Political Process Theory proposes that social movements cannot be studied in isolation from the political context in which they are embedded, since the political opportunities in those contexts are key determinants of their success or failure (see Tarrow (1998), p.27). For Tarrow, the foundational work in this tradition was Tilly's *From Mobilisation to Revolution*, in which Tilly proposed that the opportunity/threat to the activist and the facilitation/repression from the authorities where among the most important factors to predict the possible emergence and success of a social movement.

However, with the advent of new ways of communication at the end of the twentieth century and beginning of the new millennium, social movements are facing important changes in the way they organise and the reach they can have, posing a big challenge to the explicative capabilities of the previously reviewed models. Bennett (see Bennett (2003), pp. 150-163) argued that digital media changed activism in several ways:

- Loosening the connections between members of the social movement. Internet networking allows for remote and very impersonal relations. These networks are also characterised by their low density in connections. Members of Internet mediated social movements have connections with a very low number of other activists, since a dense network is not necessary to participate or to be functional inside the social movement (Van Laer and Van Aelst 2010).

- Weakening the identification of the local activists with the movement. The ease of access to a large array of causes and social movements allows individuals to identify and participate in a variety of campaigns at once, but with lower levels of personal identification with each one of them as the personal resources are more thinly spread (Van Laer and Van Aelst 2010).

- Reducing the influence of ideology on social movements. The presence of many campaigns trying to create networks with each other makes the ideological basis of such campaigns thin. In order to create the network, compromises must be made, and campaigns often prefer to lose the ideological depth of their claims so they can be associated with other campaigns and social movements. Additionally, the very low cost of networking allows for the connection of ideas that before were thought to be impossible, undermining the capacity for creating core ideologies.

- Reducing the influence of resource-rich organisations in the support of social movements. The relative low cost of Internet campaigns, the ease of creating links between organisations, the variety of channels by which organisations can reach their audiences and the unpredictability of traffic patterns may help to raise social movements from relative obscurity without the help or intervention of other more powerful organisations.

- Allowing the creation of very long term campaigns (e.g., climate change, anti-globalisation). By relying less on NGOs, unions or environmental organisations, and moving their organisational functions to volunteers or single individuals, campaigns could last much longer than in the pre-Internet era. Additionally, Bennett (2003) proposed that relying less on centralised organisation means a more diffuse control over the campaign, making it more difficult to turn off or on by targeting its central organisational support. Finally, Bennett (2003) proposed that since individuals in late modern societies are less identified with centralised political organisations, permanently rolling campaigns provide the mobilisation structures for those individuals.

With the growing penetration of digital communication in our societies, the above mentioned points have only intensified in recent years. Today, social movements recruit adherents almost exclusively by digital means —especially among younger generations—. While this facilitates almost instant communication and means of coordination, many social movements today struggle with commitments issues (see Bennett (2003); Tilly (2004), p. 110). Instant communication

also transformed the response time of social movements. Massive demonstrations can be organised in a matter of days and can disrupt entire cities and then disband and disappear with the same celerity.

Because of this, there is a growing number of social scientists who believe that we are facing a deep transformation in the way social movement and collective actions are organised and expressed (Van Laer and Van Aelst 2010). While some believe these changes are somehow an intensification of the characteristics of what defines a social movement —e.g., the number of participants, the organisation, the intensity of their actions— others believe that we are facing a new kind of social movement (see Castells (2015)), and the current models and methodologies may fall short in fully understanding this phenomenon.

This new breed of social movements is characterised by Spanish sociologist Manuel Castells (Castells 2015) as an effort of horizontalisation of social, political and economic structures. This is enabled by the Internet and social media in which horizontal communication is preferred or even encouraged by the social media platform's design. In his view, social movements in the Internet era are characterised by a leadership stemming from young, well-educated and digitally savvy individuals. Castells argued that the access to these digital networks increases the autonomy, environmental and social consciousness and self identification as global citizens of individuals (see Castells (2010), p.390), allowing them to overturn, or at least bypass, common social hierarchies, transforming the digital networks into democratisation tools. Castells' major work on social movements and digital communications is *Networks of Outrage and Hope*. It was published in 2012, after the events of the Spanish movement Indignados and the Arab Spring, in which social media played a major role in the transmission of ideas and the organisation of the social movement, allowing citizens to broadcast their demands at a totally different scale and to even overthrow existing regimes —with varied levels of success in the mid and long term— (Khondker 2011; Grinin et al. 2019). With these examples, Castell's optimism about Internet technologies was almost justified. However, his optimism may have been more moderate today with knowledge of the 2016 US presidential election and Brexit scandals, involving Facebook and Cambridge Analytica, and the rise of populist pro-capitalism and nationalistic movements in Europe and the US. Brym et al. (2018), argue that digital technologies are a double edged sword that can cut both ways – granted, Internet and social media can enhance the capabilities of individuals to create a more horizontal and democratic society, but these technologies are also

prone to manipulation by authoritarian governments and unscrupulous corporations, to stimulate conflict in directions that can be more favourable to their objectives (Spaiser, Chadefaux, et al. 2017). It is, in a sense, a new battleground for political struggles in which the victories are, again, reserved for the more resourceful players (Brym et al. 2018), and these are not necessarily the idealistic young individuals portrayed by Castells (2015). I can also add that in this new battleground the relative weight of the intervening factors predicting social movements and participation in collective actions has changed. If we follow the arguments of Tilly (2004) and Bennett (2003), digital communications reduce the relative influence of identity, ideology and the control of external resource-rich organisations (e.g., NGOs, foundations, political parties) over social movements. This coincides with the idea of Castells (2015) that, in the new landscape of the Internet era, the emotional component of activism and social movements has much greater relative importance than before, although ideological and structural factors still play an important role.

2.2 Definition of social movements and collective action

2.2.1 Social movements

Before going any further we must take some time to define core concepts. Sociologist Sidney Tarrow and social psychologist Bert Klandermans offer a compact definition of social movements. They propose that "Social movements are collective challenges by people with common purposes and solidarity in sustained interactions with elites and authorities" (see Klandermans and Stekelenburg (2013), p.2). Klandermans then elaborates upon this, arguing that this definition includes three key elements:

1. The "collective challenges," which are disruptive, collective and direct actions against authorities, elites or cultural codes (this is equivalent to the collective actions repertoire of Tilly (2004)).

2. People with common purpose and solidarity. For Klandermans these common purposes and claims have their roots in feelings of collective identity and solidarity.

3. Sustained collective action turns isolated incidents of social unrest into a social movement; or in words of Tilly (2004), a campaign.

2.2.2 Collective actions

Olson, in his 1965 book *Logic of Collective Action* proposed an explicative model of why participants would like to engage in collective actions. His model emphasised the role of rational thinking in individuals, arguing that rational actors will only take part in collective actions when selective incentives persuade them to do so. In this sense, collective action is seen as the actions taken by a group in order to acquire a collective good, understanding this as "any good such that, if a person X_i in a group $[X_1, \ldots X_n]$ consumes it, it cannot feasibly be withheld from the others in that group." (see Olson (1971), p.14). Since social movements' goals are considered public goods, the incentive for participation will decrease with the growth in participant numbers, as people will evaluate that their individual contributions will be less and less significant relative to the overall power of the social movement, motivating them to stop active participation (see Olson (1971), p.21). To counter this effect, leaders must provide "selective incentives" to convince participants that participation is worthwhile (see Tarrow (1998), p.24). This approach has been used successfully to explain why people may want to drop from participating in collective actions, but has received criticism for its capacity to explain why people would decide to participate in the first place (see Klandermans and Stekelenburg (2013), pp.774–776). Nevertheless, Olson's approximation is especially relevant for this work as it proposes the idea of a rational agent that performs logical cost-benefits calculations. However, while for Olson these calculations must be performed consciously, I will propose that such calculations can also be performed unconsciously or automatically, following a logical path, but with the intervention of emotional mechanism (Bechara, Damasio, et al. 2005; Bechara 2003).

In 2001 Wright proposed a very simple taxonomy for political actions, which also included collective actions. Wright proposed that if an individual decides to act in response to a political issue, their action can be individual or collective, dividing the collective actions into non-normative or normative. Additionally, participation can be divided in terms of its duration (ad hoc or sustained) and effort (weak vs strong) (see Wright (2003), pp.409–430). The distinctions between normative and non-normative collective actions proposed by Wright may require some attention. While one can equate the word "non-normative" to "contentious," the original wording carried much more complexity. For Wright, the normativity of the action is context dependent, but also, viewer dependent. For a member of a group, who is participating in collective actions, some actions can be absolutely essential and expected as part of the membership of such group,

and hence, very normative, while for a member of the outgroup or for a third observer, these actions can be uncommon, strange or disproportionate, and hence non-normative.

The level of violence of collective actions also plays a role in its classification. Wright acknowledged that normative collective actions tend to be more peaceful and less disruptive than non-normative collective actions (see Wright (2003), pp.409–430). This is relevant since this classification has been dominant in the socio-psychological collective actions literature from which this work draws many useful conceptualisations. Sadly, it is common to find that this fine tuned taxonomy is not properly applied —or not applied at all— in many socio-psychological publications about collective actions, rendering it almost useless. The biggest discrepancy occurs when normative and non-normative collective actions are attempted to be measured. Researchers usually use an assortment of violent and non-violent actions as examples of non-normative or normative actions, but ignoring if the listed actions are actually normative or non-normative in the specific context of the study. (e.g., Van Zomeren, Postmes, et al. (2008); Saab et al. (2015); Becker, Tausch, and Wagner (2011b); Feddes et al. (2012); Van Zomeren, Leach, et al. (2012); Van Zomeren, Spears, and Leach (2010)). This of course doesn't help in the development of a general explicative model for these kinds of phenomena, so I will move away from the normative/non-normative categorisation, and for simplicity, along the course of this work, I will classify collective actions on the following axes:

- Violent vs non-violent actions: The exact point at which a collective action may be considered violent can be subject for future studies, but in this thesis, any action that makes use of physical force or power to cause physical or psychological harm will be defined as "violent collective action."

- Low vs high cost: Relative cost can be difficult to quantify, but physical presence, meaning the physical participation of the person in the collective action (which can involve transportation), can be used to measure the cost of a collective action, in which participants will have to weigh the cost of participation and the possible outcomes much in the way proposed by Olson (Olson 1971; McAdam 1986). A long standing riot of three days inflicts a very high action cost for the participants, since it involves their physical presence in the scene, in a vulnerable and exposed situation with high stakes, for a relatively long period of time (Loveman 1998), while signing an online petition is a very low cost action, since it takes minutes, can be done from home and can be completely anonymous.

2.3 Individual psychological factors

This thesis borrows heavily from psychological literature, specifically, the social psychological literature on collective actions as emotions is one of the main predictors in the proposed analysis and models of the thesis. This being the case, a deeper dive into the main theoretical models offered by social psychology regarding social movements and activism is important. But before that, it is necessary to delimit some basic psychological concepts, which later will be key pieces in the models.

2.3.1 Emotions

A common misconception about emotions is that they are illogical processes. While it is true that emotions, in a strict sense, are irrational, as they occur before any cognitive evaluation, this doesn't mean that emotions are completely illogical. Emotion and cognition are two systems with different goals. In the case of emotions, they can protect us and help us to perceive the world around us. They evolved to produce some adaptive advantage, and almost all of them are very logical. The problem with them is their simplicity and pervasiveness, which is not always adaptive to the more complex conditions of society. Additionally to this, emotions play an important role in the motivational process which is key to understand the participation of individuals in activism and social movements, and to understand why certain social movements radicalise themselves.

Neuro-anatomically, emotions originate and are processed in the limbic system. The amygdala, which is an almond shaped structure in the limbic system, plays a major role in the modulation of emotions, but also in memory and decision making process. The limbic system is one the first regions of the brain to activate in response to a stimulus. This is facilitated by its location, around the thalamus, a region of grey matter in the midbrain which works as a sensory relay for both proprioceptors and exteroceptors (Dalgleish 2004). This means that every sensation and stimulus experienced by the human body goes through the limbic system, and is preprocessed here before reaching cortical regions for cognitive elaboration. One of the important implications of this interconnection is that it can lead to emotionally biassed the decision making process. To address this Damasio (1998) proposed the Somatic Marker Hypothesis, which states that subjects weigh the incentive of the options when making decisions, using cognitive and emotional resources. When the situation is ambiguous or the available information is limited by contextual

factors (e.g., while facing danger, threat, or under time constraints), unconscious "emotional markers" influence the final decision (Damasio 1998; Bechara, Damasio, et al. 2005; Bechara and Damasio 2005; Bechara 2003). This hypothesis provides us with a framework to understand one of the possible mechanisms by which emotions influence decision making.

Emotional learning, on the other hand, helps to understand why certain memories are more strongly related to emotional reaction than others. Emotional learning is strongly associated with fear, phobias and the avoidance of harm. It is very strong and pervasive, and sometimes requires no more than one exposure to the experience to be acquired, with life-long consequences. In humans, fear learning has been associated with anxiety disorders, depression and PTSD (Flor and Birbaumer 2015; Fanselow 1990; Delgado et al. 2006).

Now, the traditional neuro-psychological conceptualisation of emotions proposes that they are first "acted out" in the body and later felt by the individual (James 1884; Ellsworth 1994). Under this characterisation, a subject facing a threat to their life first has all the autonomic body reactions preparing it to "flee or fight" and only later does the subject experience fear. This process occurs very fast and unconsciously, which explains why the final experience of fear seems to be concurrent with the body reactions. The first major proponent of this conceptualisation of emotions was James in his paper "What is an emotion?" (James 1884). This idea was later expanded by Damasio in his hypotheses of the generation of the consciousness from emotional reactions (Damasio 2000), and the Somatic Marker Hypothesis (Damasio 1998; Verweij and Damasio 2019; Bechara and Damasio 2005; Bechara, Damasio, et al. 2005). Damasio provided substantial evidence supporting the idea that neuro-physiological and body reactions to an emotional stimulus come before cognitive elaboration in neocortical regions of the brain, and hence, before the conscious perception of the emotion (Damasio 2000; Parvizi and Damasio 2001; Parvizi, Van Hoesen, et al. 2006; Bosse et al. 2008).

To Damasio emotions are unconscious neural reactions triggered by specific stimuli, which take place in complex neural circuits in the brain. These circuits have the predisposition to create specific somatic and behavioural reactions in their subjects, which can later be observed by external agents or by the individual itself (Bosse et al. 2008). A clear distinction exists in the work of Damasio between emotions and feelings. For Damasio, a feeling is the posterior mental image plus cognitive and behavioural responses triggered by the initial activation of the emotional circuits. This feeling is still an unconscious phenomena which must be later

processed by a second order cortical structure. The realisation of "feeling a feeling", or the phenomenological awareness of experiencing a feeling, is what creates consciousness (Bosse et al. 2008).

In Damasio's work some pre-wired neural circuits can be identified for fear, anger, sadness and happiness (see Damasio (2000), pp.60–61; Bosse et al. (2008)). These four emotions were never proposed as universal basic emotions by Damasio. The notion of "basic emotions" was actually popularised by Ekman and Friesen (1971) and later expanded in Ekman (1999) and in Dalgleish (2004) in which he proposes suggests six basic emotions: happiness, sadness, fear, anger, disgust and surprise. Ekman's work have been hugely influential in the field of computer-aided emotion recognition, and is today probably the most commonly used emotion taxonomy (for examples see: Li and Deng (2020) and Binali et al. (2010)); And although the idea of universal "basic emotions" has been contested (Jack, Garrod, Yu, et al. 2012; Mesquita et al. 2016; Heyes 2019), a clear taxonomy of emotions is very useful for the tasks of emotion recognition and emotion classification.

However, recent research in cross-cultural emotion recognition suggests that in order to have a more valid cross-cultural taxonomy of basic emotions, the range of those emotions should actually be narrower, including ideally only four emotions: anger, fear, sadness, and joy (Jack, Garrod, and Schyns 2014; Jack, Sun, et al. 2016). These findings in cross-cultural emotion recognition are also supported in recent research on emotion expression in animals (Gu et al. 2019), where the same four emotions are suggested to be "hard-wired" in the neural circuits of the drosophila fly.

It is worth noting that the above conceptualisation of emotions is not the only one. Appraisal Theory of Emotions (Arnold 1960) has been a popular framework to understand emotional expression in humans. Appraisal theory proposes that emotions and their differentiation are caused by an appraisal of a stimulus i.e., its fit or mismatch with goals, values, expectations and other cognitive components. In this work, I argue that at least some basic emotions (fear, anger, sadness and joy) are biologically rooted and hence do not need an appraisal to emerge as distinct emotions (Ekman 1992; Damasio et al. 2000; Dalgleish 2004). This does not mean, however, that these same emotions cannot be modulated later by cognitive processes, or that other more complex emotions could not emerge when this modulation/appraisal happens (Ekman 1992; Jack, Garrod, and Schyns 2014). However, neurological evidence suggests that there are

emotions that are biologically rooted, and hence do not need cognitive appraisal to emerge (Jack, Garrod, and Schyns 2014).

There seems to be a growing consensus that for a widely valid taxonomy of emotions, a simpler model of four "basic emotions," namely anger, fear, sadness and joy (Damasio et al. 2000; Dalgleish 2004; Jack, Garrod, Yu, et al. 2012; Jack, Garrod, and Schyns 2014; Jack, Sun, et al. 2016), should be used. Specifically, I will use Damasio's taxonomy and conceptualisation of the four emotions anger, fear, sadness and joy (see Damasio (2000), pp.60–61; Bosse et al. (2008)). These four emotions have their distinctive neurological circuits and this is why they are highly distinguishable from each other with fairly standard expressions from subject to subject (i.e., humans are "hardwired" to detect them and produce them, see Jack, Garrod, and Schyns (2014)). This makes this taxonomy ideal for classification problems like the one encountered in this project, as a small set of fairly distinguishable categories has a significantly better chance of producing accurate classification than a more comprehensive but overlapping set of emotions. In this project I am favouring precision of classification over comprehensiveness, and this is why the set of four emotions fear, anger, sadness and joy will be measured in the collected Twitter data.

Finally, understanding the origin of emotions, and having a clear taxonomy for them, is relevant for the study of social movements as it allows to understands the the motivational processes that lead to activism (see Marcus et al. (2022); Arzheimer et al. (2017), pp.406–425). In Damasio's work we are led to conclude that emotions can act unconsciously and before other modulatory cognitive processes over our reaction to external or internal stimuli. This means that the emotional bias over our actions and decisions is almost biologically inevitable, although it can be later modulated by rational deliberation. However, if emotions can also affect other automatic cognitive processes, such as, for example, selective attention, its bias effect could be more pervasive, even to our best conscious efforts to counter them. This makes the understanding of emotion and its effects on other cognitive processes very relevant to understand social phenomena, especially in the cases in which such phenomena are charged with emotional expressions like social movements.

2.3.2 Motivation

As we saw previously, emotions are the consequence of a series of neuro-biological reactions which can lead to cognitive, somatic and behavioural outcomes in individuals. However, emotional states are not enough to generate outcomes in the subjects. This is when motivation, as a process, can help us to understand the more complex dynamics of human behaviour.

Depending on the field of study, the definition of motivation can have very different emphasis, from focusing on its neuro-biological basis (Simpson and Balsam 2016; Berridge 2004; Hughes and Zaki 2015) or its psychological basis and its correlates with self, identity, learning and goal oriented behaviour (Bargh et al. 2010; Deci and Ryan 2010; Swann and Bosson 2010). During the course of this work, I will understand motivation as the reason "why a person in a given situation selects one response over another or makes a given response with great energization or frequency." (Bargh et al. 2010), understanding "energization" as the level vigour or fervour a person puts into its actions. I believe this definition is general enough for the purposes of this thesis.

Neuro-biologically, a number of different factors influence the motivational process in individuals. These include internal physiological states (such as health, stress, circadian cycle, nutritional needs), environmental factors (presence of an opportunity, the presence of a harm, the conditions to reach a goal, etc.) and the past history of the individual, which can be represented in the form of memories, knowledge, previous conditioning and other previous cognitions. These factors must then be weighted by the subject to determine their costs and benefits, to finally generate the behavioural response (Simpson and Balsam 2016).

Internal physiological factors usually deal with keeping the homeostasis of the body at the optimum. For example, in cases when proprioceptors detect a deficit of water in the body, thirst and the desire to drink water are triggered in order to correct this deficit. Similar mechanisms are involved in hunger, body temperature maintenance and the craving of very specific nutrients such salt or calcium. They are also present on more complex behaviours such as some manifestations of sex desire and aggression (Berridge 2004). These motivational drives do not present large variations among different subjects as they are rooted deep in the biology of the human body and its cost-benefit outcomes vary within narrow intervals. Drives related to the environment and to the past history of the subject, however, can be affected by intervening variables. The cost-benefit calculations regarding environmental and historical factors are not

as automatic as those involving homeostatic factors. Previous knowledge, memories, attention and psychological needs can interfere in the outcome, giving space for more variability. Considering also that other motivational factors can be present at any given moment in most of the complex social situations in which individuals are embedded every day, achieving balance between these competing motivators will be much more dependent on individual characteristics (Simpson and Balsam 2016). To add even more complexity, there is evidence suggesting that motivation can highly influence the acquisition of memories and knowledge, in a process known as "motivated cognition". Motivated cognition can influence perception, attention and decision making, generating bias (like confirmation bias and in-group bias) in an effortless and pervasive manner (Hughes and Zaki 2015). All this means that biassed acquired memories can perfectly affect motivation for a different action in future processes of decision making.

Finally, there is growing evidence suggesting that motivational processes can be influenced by emotions in a variety of ways. Emotions have been shown to influence the intervening processes of motivation, such as perception, attention, learning, memory retrieval and decision making (Sanfey et al. 2003; LeBlanc et al. 2015; Levine and Safer 2002). This supports the idea of Damasio that emotions are a fundamental part in the cognitive and social processes of individuals, and that without them, normal social functioning is almost impossible (Damasio et al. 2000).

Now, the previous neuro-biological notions, while very useful to understand basic brain functions, fall short to conceptualise higher-order processes in human beings. Social psychologists have therefore moved the focus of the study of motivations to the concept of goals. If we want to follow the computational-cognitive metaphor of the functioning of the human brain, we can think of a goal as a higher order abstraction of the basic brain mechanism described by neuro-biology and neuro-psychology. Goals are internal and subjective states and processes, formed by more basic cognitive structures and mediated by neuro-biological processes (Bargh et al. 2010). This contrasts with the idea of a "stimuli", which, in a sense, is always external even if it is a perception of the own body. Goals are also more complex cognitive structures, involving processes of ideation, memory and attention. This is what allows the creation and pursuit of a goal. This goal creation and pursuit can occur consciously and unconsciously. Conscious goal setting and pursuit opens the door for the idea of self-regulated behaviour and motivation. Planning and expectations are consequences of conscious goal setting as well as the expectations

and ideations regarding the consequences of achieving those goals. This mechanism helps to keep individuals engaged in goal pursuit and is part of the many mechanisms helping keep any goal-related motivation (Bargh et al. 2010). It can be argued that the conscious goal pursuit is intentional, as the self controls the creation of the goals and the plans and actions to achieve them. This implies that the self governs the brain and inhabits it. This feeds into the ideation of agency and agency beliefs such as personal efficacy and self-efficacy (see Ford (1992), pp.123–130; Reeve (2018), pp.226–237), which are crucial for the creation of memories, which will later influence motivational processes.

There is also the possibility for unconscious goal setting and pursuit, which can bypass completely the awareness of the subject. Psychologists have tried to conceptualise this phenomenon through cognitive perspectives —in which the unconscious perception of a stimulus can trigger the generation of secondary unconscious goals in the mind (such as self-esteem and self protection goals)—, or under a more psychoanalytic perspectives —in which the unconscious section of the ego, superego and id can influences the goal and goal pursuit strategies— (see Reeve (2018), p.383; Bargh et al. (2010)). This is very interesting as it allows us to think of situations in which unconscious motivations and goals, that may not be socially acceptable, are masked by conscious and explicit motivations, allowing for example subjects with clear racist goals to believe to have non-racist motivations as their goals.

In any case, goal setting and goal pursuit seems to be very related with the self and the protection of the self, which in turn is related to the construction and protection of the identity of individuals (see Reeve (2018), p.274)

Finally, understanding the process of motivation help us, first, to understand how emotions play a role in guiding motivational processes, by conscious and unconscious mechanism, but also help us understand how the “goal setting” and “goal pursuit” supra-processes can shape the creation of the self and identity, which are both mental structures at the centre of the motivational processes leading to participation in social activities, and ultimately, in the participation of collective actions and social movements.

2.3.3 Self and Identity

The concept of self in popular culture has been treated in many different and often contradictory ways. On the one hand there is a certain popular psychological literature that promotes the

enhancement of the self (Csikszentmihalyi and Seligman 2000), arguing for the creation of what can be called "egocentric" goals and ways of life, supporting this on the basis of values of independence and freedom. Some others have proposed that the self is a kind of burden that must be left behind in order to acquire peace and wisdom, as if the self is something that can be extirpated from the mind (Leary 2004). Here, I will take distance from these approximations to the concept of self to instead approach to a more Jamesian conceptualisation. For me, as it was for James et al. (1890), p. 336, the self is a psychic structure that provides "connectedness" and "unbrokenness" to individuals. In this sense, the self cannot be removed or ignored, neither can it be enhanced, as it is a structure always present and vital for the correct functioning of the mind. This structure is built upon a set of representations about oneself, parallel to representations that we have about other individuals (Swann and Bosson 2010). The absence of an integrated self will mean serious maladaptive disorder in the individuals (Sass and Parnas 2003). The idea of the self as a functional structure also opens the door to think about the self as a dynamic structure or for the existence of multiple selves. In order to understand the articulation of an adaptive self we must introduce the idea of self-concepts. Self-concepts are cognitive structures containing content, attitudes or evaluative judgements and are used to make sense of our surroundings, set goals and protect our self-esteem (see Leary and Tangney (2011), p.72). Using this concept we can imagine the self as the background process that *thinks* about the individual. The content of those thoughts are in part what we might call the self-concept.

These self-concepts are flexible and adjustable, and many of these structures can exist in order to help the self to adapt to multiple situations (e.g., the self-concepts of a person as a parent is different to their self concept as a citizen). Additionally, self-concepts tend to organise around some relevant aspect of the life of individuals (gender, ethnicity, age, educational attainment, etc). The degree of organisation of these self-concepts can be related with the ability to process the information associated with these domains. Better organised self-concepts can process and classify information more efficiently, make better decisions and identify more achievable goals (see Leary and Tangney (2011), p.73). An important element of self-concepts and their organisation is that individuals tend to guide their decisions in order to protect and improve their own evaluations about their self-concepts. For example, if I consider myself a good citizen, my actions and goals will be guided by this concept in order to be maintained. Also, if I consider myself as part of an ethnic minority, my efforts will be guided to improve my evaluation about me being part of that ethnic minority. This has implications for the study of social movements and

collective actions. Individuals taking part in social movements usually have strong self-concepts about their membership in those movements, and their motivations to join them might be the consequence of external circumstances which threaten other important self-concepts of the same individuals.

The concept of identity in psychology is used to denote the "totality of one's self-construal" (Weinreich 1986), that is to say, the totality of memories, concepts, attitudes, values, affiliations or any other cognition that an individual considers part of themselves in the past, in the present, and how they aspire to be in the future. This identity is the psychological structure that allows the continuity of the individual's memories and gives coherence to their actions in order to construct a life story (see Erikson (1993), p.211). It connects all the individual's self-concepts in one organised and coherent supra-structure, maintaining and promoting self-concepts coherent with the supra-structure while inhibiting the incoherent ones. It also plays a role in the creation of goals as it allows for continuity between the past, the present and the future. Probably one of the most notable characteristics of self-identity is its ability to appear as a stable and unique structure across different situations and along the life of the individual. However, this is not the case, and individuals often underestimate how much their identity, self-concepts and goals will change in the future, regardless of their age (Quoidbach et al. 2013). Moreover, an individual can have multiple identities depending on which social situation they are facing (see Bargh et al. (2010); Leary and Tangney (2011), p.73). However, we unconsciously create these identities to be related and coherent with each other. Otherwise, there is a risk of crossing the border to psychological disorders.

Identity can have multiple aspects. For example, ethnic and gender identity are known to be very important parts of individuals' lives (Phinney 1990; Egan and Perry 2001), to the point that their denial by others members of society can lead to serious psychological consequences. However, there is some aspect of identity which is especially important for our purposes, which is the one that has to do with group affiliations and social interactions. This is known as Social Identity and is key to understanding collective actions and social movements from a socio-psychological perspective, as well as being the base for all modern socio-psychological models of collective actions.

2.3.4 The Social Identity Theory

Social Identity Theory (SIT) was first proposed by Henri Tajfel and John Turner in 1979 in their work *An Integrative Theory of Social Conflict* and later refined in 1986 with *The Social Identity Theory of Intergroup Behaviour*. It introduces the concept of "social identity" as a part of the self-image of individuals. This social identity emerges from the membership in some relevant social category or group (Tajfel and Turner 2004; Tajfel, Turner, et al. 1979), together with the value and emotional significance attached to that membership (Tajfel 1978).

From Tajfel and Turner's work we can infer that social identity has, at least, three main components: a cognitive one (the awareness of membership); an evaluative one (the value the individual gives to the membership), and an emotional one (feelings and affections towards the other members of the group). One can also infer that the social identity of an individual is composed of several group memberships. For example, an individual can be Catholic, a student, a member of a football club, and a sympathiser for a political party. The conjunction of these group memberships creates the individual's social identity, and each of these memberships can have different degrees of cognitive significance, emotional involvement and overall value for the individual. It is important to note that social identity must not be confused with collective identity. The former is related to cognition of an individual about its membership in one or more groups (Tajfel, Turner, et al. 1979). The latter is defined as the cognitions shared by the members of a single group and is related to the phenomena of self-categorisation (the phenomena by which an individual categorises itself in a group). The Social Identity Theory was formulated to understand intergroup behaviours (Tajfel and Turner 2004; Tajfel, Turner, et al. 1979) —in particular, prejudice and ingroup bias— and not as a theory of social categorisation. It does not provide many tools to understand why or how people end up being categorised into certain social groups; instead, it provides a framework to understand the collective behaviour of individuals who already belong and self-categorise themselves as members of a group.

Tajfel and Turner proposed three basic assumptions on which the SIT rests:

- Individuals strive to maintain or enhance their self-esteem,
- Social groups and categories can have positive or negative value connotations. Hence, the social identity of an individual can also be positive or negative depending on how much the memberships in those social groups and categories contribute to it,

- The evaluation of one's group is done in reference to another group through social comparison (Tajfel, Turner, et al. 1979).

These assumptions lead to a couple of very simple but important corollaries. Individuals will strive to maintain or improve their social identity, and, if the result of the social comparisons leave them with unsatisfactory results, individuals will try to leave their un-prestigious groups, or, in the case where this is not possible, they will try to improve the prestige and social evaluation of their groups (Tajfel, Turner, et al. 1979).

Now, assuming that for some individuals changing their group memberships can be very psychologically and socially costly —which is often the case with most of the relevant social categories e.g., nationality, religion, gender—, and that the individuals subjectively identify with the relevant groups or categories in a way that makes it part of their self-concept, two possible scenarios are proposed. The first one is the process of differentiation of the ingroup from the outgroup (Tajfel 1974). This process has the objective of maintaining or achieving superiority over the outgroup in some relevant dimension, and is most likely to occur in situations in which the status of the groups is highly stable and the legitimacy of the status differences is not under question. Eventually, this avenue leads to the creation of intergroup competition, rivalries, conflict, and eventual discrimination and prejudice. On the other hand, when the status of the groups is unstable —meaning that the prestige of the group can be improved— and this status is perceived as illegitimate, collective action and social unrest are the most likely outcome (Tajfel 1974).

In sum, according to the Social Identity Theory, individuals engage in collective actions to improve the status of their group if they cannot leave it, if they perceive that the status can be changed, and finally if they perceive that the status differences are illegitimate (Van Stekelenburg and Klandermans 2017). Additionally, the struggle of the ingroup elicit emotions on the ingroup members, since as Social Identity Theory suggests, the group identity is a very real part of the individuals identity. This emotions are come to be known as inter-group emotions (Mackie et al. 2008), being this emotions felt by individuals, but that relate to the group-identity of them. A similar concept is "emotional climate" (De Rivera 1992), which refers to the emotions/feelings between members of the groups, and not regarding events affecting the group. This concept describe a necessarily more stable phenomena, and hence is more appropriate to describe the basal state of groups or social movements, instead of their immediate reaction to events.

Now, Social Identity Theory does not necessarily apply to all social movements. A very good example of this are the new environmental and peace seeking social movements. These movements seek to set an agenda and promote a topic that is deemed to be important for everyone. Nevertheless, reviewing the Social Identity Theory is relevant in order to understand more contemporary models.

2.4 Socio-psychological models of collective action

In this section I will review some models and studies that have been most influential in the conceptualisation of this work, emphasising the empirical evidence supporting these models as well as their relationship with the emotional state of the participants.

2.4.1 SIMCA: the Social Identity Model of Collective Action

In 2008, van Zomeren, Postmes and Spears published an impressive meta-analytic paper reviewing more than 180 individual effects of perceived injustice, efficacy, and identity (in their socio-psychological connotation) on collective action, in an effort to construct an integrative predictive models of activism. The compiled dataset contained a total of 27 individual groups (27 studies) and more than 3000 data-points, which allowed them to obtain significant statistical power. They showed that in isolation, the three predictors mentioned above have a moderate predictive capability on the intention of participation in collective actions. They showed that emotionally charged predictors, such as affective injustice and politicised identity produced stronger effects than the non-emotionally charged counterparts. Additionally, they were able to establish that identity was the most distal predictor, suggesting that the motivation for participating in activism originates in the social identity of the participants but is mediated by the perceptions of injustice and efficacy, which acted as proximal predictors of collective actions. Finally, they also were able to isolate the individual effects of the three main predictors concluding that all of them have unique predictive capabilities on the dependent variable even when controlling for between-predictor co-variance (Van Zomeren, Postmes, et al. 2008).

In order to test more directly their model, Van Zomeren, Postmes, et al. (2008) performed a path analysis which allowed them to arrive at the model presented in Figure 2.1.

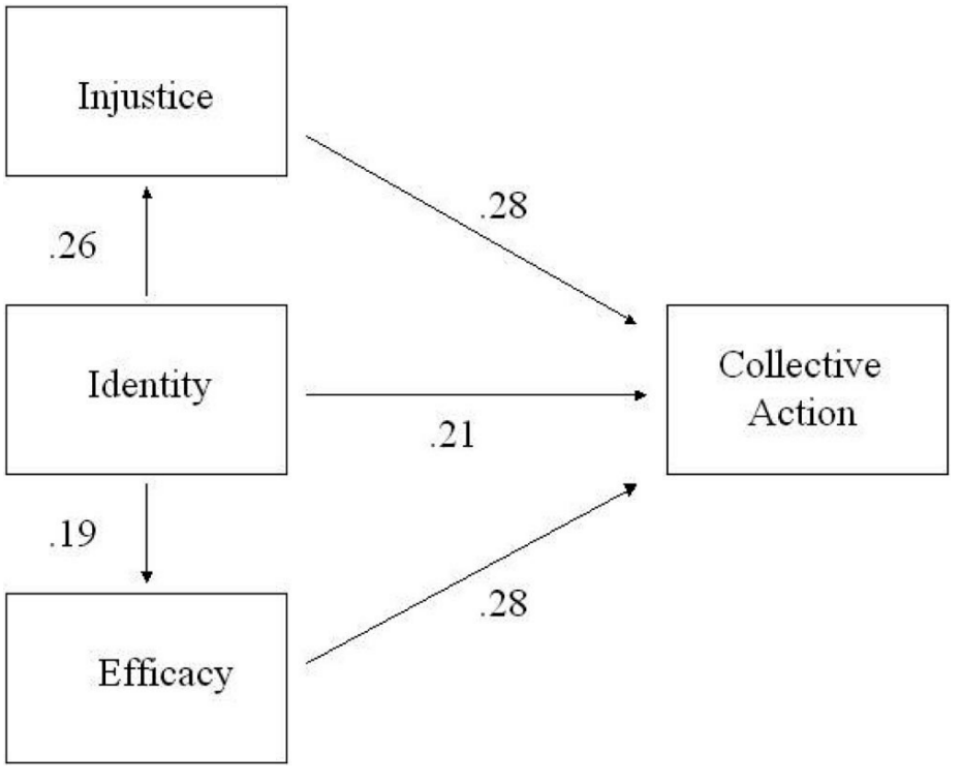

Figure 2.1: Social Identity model of collective action (Van Zomeren, Postmes, et al. 2008)

In Figure 2.1, we can see that Identity is moderately related to the intention of participating in collective actions and has a moderate relationship with the other predictors. We can also see that the perception of injustice and the perception of group efficacy have a greater direct effect on the intention of participating in collective actions. If we consider that these effects are independent of the covariance between the predictors, this models provides important evidence suggesting that a strong identification with the ingroup is not a necessary condition for participation in collective actions, but instead, is one of the possible sources, or distal predictor, of other variables that are more closely related to the participation in collective actions.

In sum, this meta-study provided robust evidence that certain individual factors related to the identity of the individual and the perception of injustices as well as of self-efficacy were reliably related to the participation in collective actions. However, this model does not yet differentiate the types of emotions involved in this process, although it is implicitly suggested that they may have to be present as reactions to perceived injustices.

2.4.2 The Dual Pathway Model

The development of the model suggested by van Zomeren and his group continued in the coming years, exploring the possibility of a more integrative and fine tuned model to better predict the engagement of individuals in collective actions. In 2012 they published *Protester as "Passionate economist": A Dynamic Dual Pathway Model of Approach Coping with Collective Disadvantage* (Van Zomeren, Leach, et al. 2012). In this paper, the authors proposed that collective actions are a result of a coping mechanism for disadvantaged groups. This coping mechanism describes two distinct processes: an emotion-focused and a problem-focused approach to coping. We can see here that the inclusion of an affective component as part of the main model now becomes explicit in the work of van Zomeren, giving it the status of one of the two mechanisms through which the individuals can engage in collective actions. Under this model the emotion-focused approach revolves around the experience of group-based anger, which is the result of the perception of unfair collective disadvantages, while the problem-focused approach revolves around the experience of group efficacy, which is the result of the perception of a reasonable level of group coping potential for social change. In other words, one mechanism is based on emotional motivations which are the product of a perceived wrongness, and the other is based on a more instrumental calculation of the probability of success.

A key element of this model is the continuous reappraisal of the disadvantageous situation of the ingroup along the execution of collective actions (see Figure 2.2), which, it is hypothesised, can inspire future engagement in collective actions depending upon the outcomes.

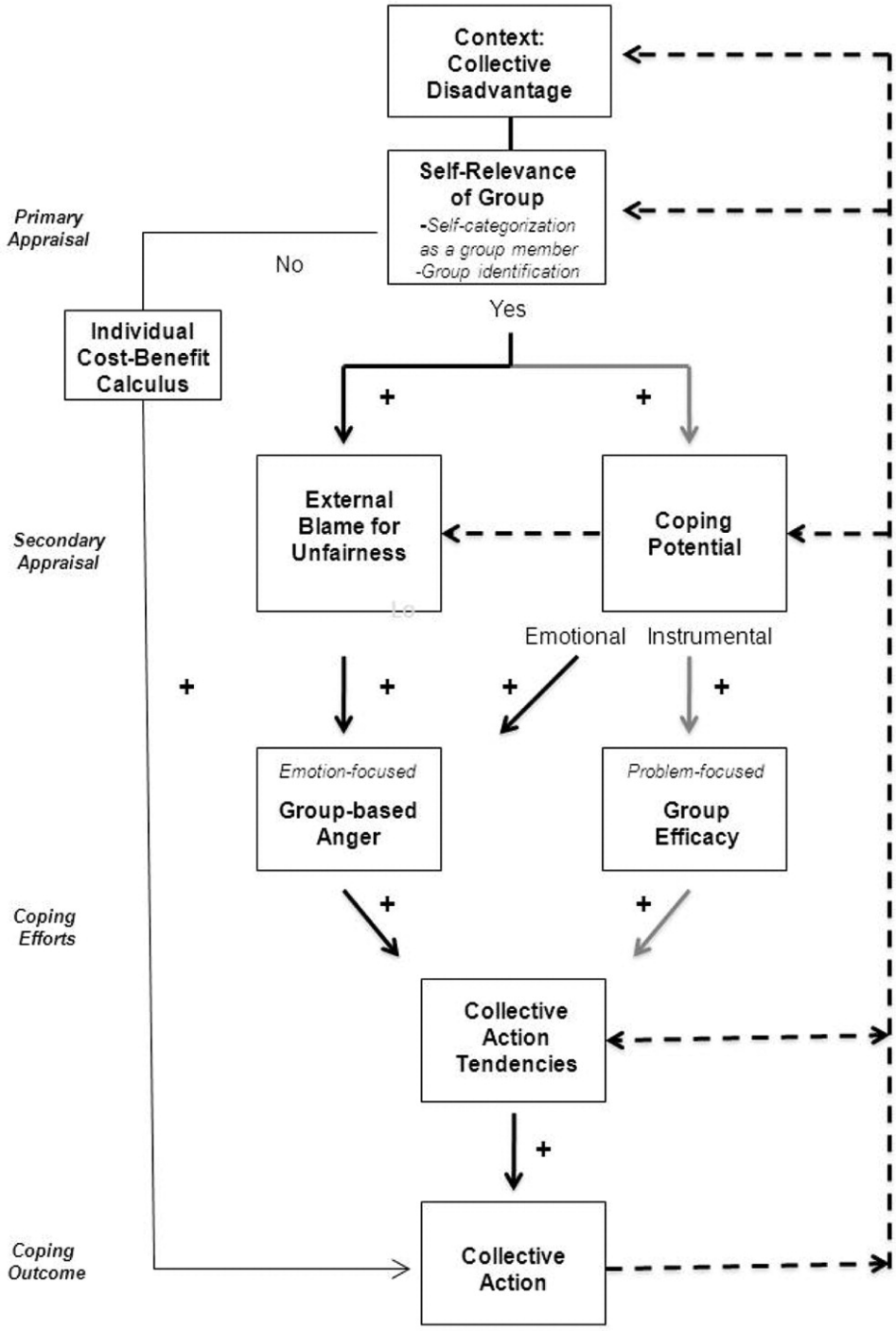

Figure 2.2: The dynamic dual pathway model of coping with collective disadvantage (Van Zomeren, Leach, et al. 2012)

This model has been partially supported by evidence (Van Zomeren, Postmes, et al. 2008; Van Zomeren, Spears, Fischer, et al. 2004; Van Zomeren, Spears, and Leach 2010), however its feedback mechanism is more challenging to test reliably as it implies the testing of the emotional reactions and attitudinal changes of individuals after the actual participation in collective action or during a campaign organised by a social movement. This would require the

use of panel studies, which adds significant complexity to the data collection and study design. Additionally, the modelling strategies required to test such hypotheses grow in complexity and decrease in accuracy and parsimony as more variables are included in the model. In this case, an appropriate technique could be the use of cross-lagged structural equations. These techniques could, theoretically, provide support for the model, provided that the relationships between the variables across time are strong enough to be detected (Burkholder and Harlow 2003; Pakpahan et al. 2017; Kenny 2014). To my knowledge, the full theoretical model hasn't been tested yet.

What is especially relevant for this work is the acknowledgement of the role emotions can play as a motivator for and a trigger of collective actions. In this same paper Van Zomeren, Leach, et al. (2012) proposed the idea that perception of efficacy of the collective actions can increase the feeling of empowerment of the participants, which can lead to an increased perception of unfairness and an increased level of group-based anger as a result (Van Zomeren, Leach, et al. 2012). Finally, the authors suggested the idea that other emotions can also play a role in motivating collective actions, citing a study in which fear was found to be a proximal predictor of intention of participation in collective action against climate change (Van Zomeren, Spears, and Leach 2010).

2.4.3 ESIM: Elaborated Social Identity Model

In parallel to the work of Van Zomeren, Spears, Fischer, et al. (2004) other social psychologists were developing a similar model to understand participation in collective action from the individual's perspective.

The elaborated social identity model (ESIM) (Drury and Reicher 1999; Reicher 1996; Stott and Drury 1999; Stott and Reicher 1998) is essentially a psychological model of crowd behaviour. In this sense, it deals with collective actions as the one phenomenon of interest, whereas the previous models place collective actions as the outcomes of the models, and in many cases don't deal with the development of the collective actions nor with their consequences. The ESIM can only deal with certain types of collective actions, namely the ones that involve groups of people in crowds (demonstrations, strikes or riots). Its main concern is the appearance of a feedback loop in crowds, which enables them to empower themselves. This feedback loop occurs when the participants of the collective action perceive, on the one hand, that they are part of a larger group which shares their concerns, and on the other, that as a group they can have more power

to pressure for change. These perceptions, in turn, boost the feeling of collective empowerment of participants in collective actions, which is defined as "the perceived degree of control that members of one group have over their fate and that of the other groups." (Drury and Reicher 1999). This feeling of empowerment will be triggered by an intergroup event in which one of the groups experiences grievance, causing its unification from an initially disunited set of subgroups in a crowd (Drury and Reicher 1999). This unification leads to a sense of common fate, which eventually motivates them to take action.

What is relevant in this model for this work is that there is evidence that the feeling of collective empowerment can also lead to a greater feeling of group-based anger (Drury and Reicher 2005). This group-based anger will be a consequence of the dis-empowerment of the activists after a clear rejection by the outgroup (government, authorities) to concede to demands they considered very legitimate. This means that emotions also have the potential for self-magnification during collective actions, which, for Drury and Reicher (2005), may be the mechanism behind the emergence of riots in crowds.

However, the Drury and Reicher (2005) models have limitations for the study of collective actions. The most important one is that the mechanisms between grievances and empowerment is not entirely clear. Additionally, grievances are seen as the main cause of group unification and the main motive for participating and organising collective actions, neglecting contextual factors that greatly influence the rise of social unrest. Many social groups around the world are currently suffering from collective grievances, but we are not seeing them all rising up. A number of other psychological and social factors are needed to create the conditions for social unrest, including —but not limited to— the amount of organisations of the oppressed group; the amount of organisation and power of the oppressing organisation they are facing; the amount of material and/or economic resources the group can gather in order to create an effective resistance, and the amount of geographical concentration of the oppressed population, *inter alia* (see Gurr (2015), pp.317–334).

2.4.4 The Van Stekelenburg & Klandermans model

In another effort to provide an explanatory framework to understand the basis of motivation for participating in collective action, Van Stekelenburg and Klandermans (2017) propose an integrative model, which includes four key components: 1) Identity, 2) Instrumental motivation,

3) Expressive motivation, and 4) Emotions.

Van Stekelenburg and Klandermans base the identity and emotional components of their model in the same socio-psychological perspective presented in this work, so a complete redefinition of them is not necessary. However the notions of instrumental and expressive motivation may need a closer look. For Van Stekelenburg and Klandermans, collective actions with strong instrumental motivation are those, which have some "external" goals like better labour conditions or access to basic civil rights. These external goals can be associated with the motivations of older social movements, with demands in terms of improving the material conditions of oppressed groups. Conversely, expressive motivations refer to the need of the social movement to express the discontent it has with something considered unfair or illegitimate. In this case, the collective action itself is the goal, with no other interests or further demands. However, Van Stekelenburg and Klandermans (2017) argued that these motivations are rarely present in a pure form in social movements, for which they proposed the use of the term of motivational pathways —one external and one expressive— contributing to a general motivational strength of the movement. In this way, both sources of motivation can coexist in the same model.

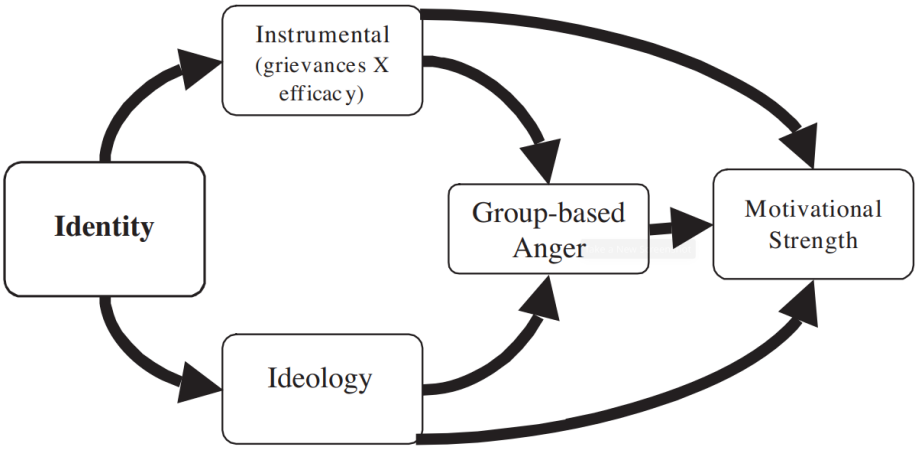

Figure 2.3: Integrative model accounting for protest motivation (Van Stekelenburg, Klandermans, and Van Dijk 2011)

In Figure 2.3 we can see how the different elements of the Van Stekelenburg and Klandermans (2017) model are integrated. It is very interesting to see again that the emotional component of the model is the most proximal predictor of the dependent variable, which in this case is the motivational strength to participate in collective actions. Also, the model states that instrumental and expressive motivation can contribute to group-based anger in addition to

their contributions to the motivational strength. In sum, this model again supports the idea proposed in this work that the emotional component plays a central role in the motivational process of collective actions. However, this model still considers only one emotion and one target of the emotion —namely, the authorities or those in power—. The model does not account for differences in types of collective actions (e.g., violent and non-violent) and does not mention how the onset of these types of collective actions could be predicted. Other research proposed a change to this, with very interesting results for the predictive capabilities of the models.

2.4.5 Different emotions leads to different actions

Until now all the reviewed models have, more or less, the same structure: Identity leads to some emotional reaction due the appraisal of unfair or illegitimate situations towards the ingroup, which then leads to the motivation to participate in collective actions. Arguably, there is some difference in the structure of the motivational process between the different models, but once they reach their emotional component, they all behave the same: we have only one emotion —usually anger or group based anger—, towards only one target —usually the authorities—, and predicting the motivation to participate in one big category of collective actions. While this is not necessarily problematic it is certainly limiting, and researchers have reckoned that the phenomena of collective actions elicit more than one important emotion (Van Zomeren, Spears, and Leach 2010). More importantly, different emotions can have different motivational outcomes. For example, fear is notoriously known for being an inhibitory emotion and anger for being a reparatory emotion. However, it is also important to note what the target of the emotions is. As Van Zomeren noted (Van Zomeren, Postmes, et al. 2008) fear of climate change can motivate the participation in collective actions, but it is likely that the participants of climate change demonstrations, while feeling fear about global warming, also feel anger towards the government for its inaction.

Responding to this problem, and following the ideas of Van Zomeren, Spears, Fischer, et al. (2004), Tausch, Becker, Spears, et al. (2011b) proposed that the motivation to engage in "non-normative" collective actions (usually violent and very disruptive actions. E.g. barricades, riots, arson, etc.) can be predicted separately from the motivation to engage in "normative" collective actions (usually non-violent actions. E.g. demonstrations, strikes, letters, etc.) by distinguishing which emotion is the one at the base of the motivation. They argue that anger —or group-based anger— is a relatively benign emotion which often seeks to restore damaged

relationships of trust (humans generally feel anger when someone or something has disappointed them in some way, but they still have and want to keep the relationship). Anger —and the presence of group efficacy— would be the main motivator of normative collective actions —or non-violent collective actions— as these actions will seek reparation of the relationship. On the other hand, Tausch, Becker, Spears, et al. (2011b) proposed that non-normative collective actions —or violent collective actions— are the result of low group efficacy, but more importantly, a new emotion: contempt towards the outgroup. Contempt does not seek the reparation of damaged relations but the destruction of it and the destruction of what is causing it. This was subsequently supported with various studies exploring slight variations of the original hypothesis and with a variety of samples (Becker and Tausch 2015; Tausch and Becker 2013; Saab et al. 2015). These results led Becker and Tausch (2015) to the development of an integrative model for the prediction of normative and non-normative collective actions, including the developments of SIMCA, a set of barriers to the emergence of collective actions, and a set of emotional and attitudinal outcomes from the collective actions. These outcomes include emotions such as joy, pride and anger directed towards the ingroup and the outgroup. The main conclusion of this model is that these outcomes can affect the motivation of the participants in future collective actions, which can lead to a loss of interest or to a process of radicalisation, depending on the particular configuration of attitudes and emotions. In Figure 2.4 we can see the full model proposed by Becker and Tausch (2015).

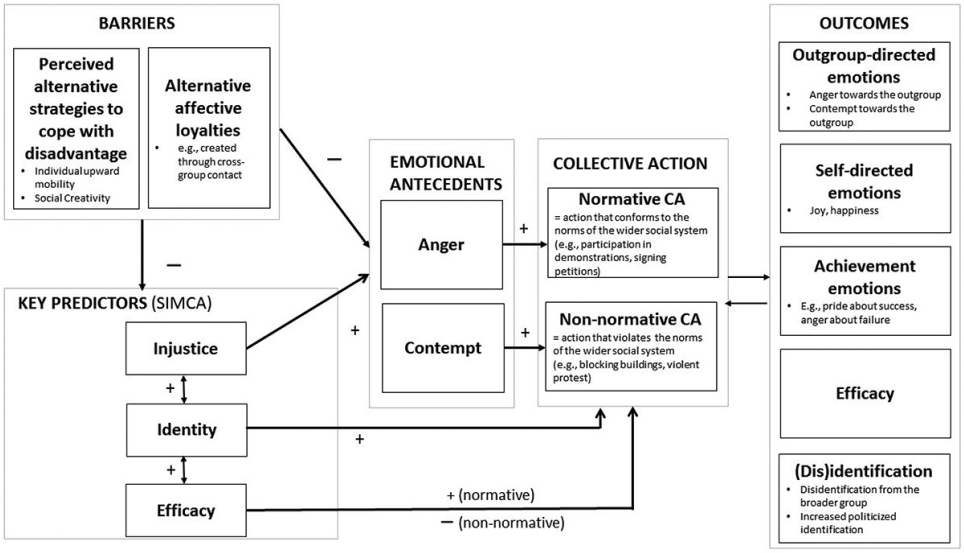

Figure 2.4: Integrative model of Becker and Tausch (2015)

What is particularly interesting in these models is the inclusion of several emotions, targeting

more than one agent, and predicting more than one type of collective actions. This model is the first that tries to address the complexity of the emotional reactions experienced by the participants in collective actions. This has paid off by allowing differential prediction of violent and non-violent collective actions, which allows the prediction of radicalisation and violent social movements.

2.5 A model to use with Twitter data

So far, I have reviewed several socio-psychological models that try to explain the motivational mechanism by which individuals participate in collective actions. While each of them put accent in different aspect of the process there is some important commonalities:

- **Emotions**: Emotions are explicitly present in nearly all of the models exposed above, with the exception of Van Zomeren, Postmes, et al. (2008) early model, although in later iterations (see Van Zomeren, Leach, et al. (2012)) the emotional component was explicitly added. In the reviewed literature, emotions are the consequence of the group-level grievances or group-level motivators, and the most common emotion associated with participation in collective actions is **anger**. Tausch, Becker, Spears, et al. (2011b) also added "contempt" with the intention to measure its differential predictive capability towards two different kinds of collective actions: Normative (mainly non-violent actions) and non-normative (mainly violent actions). In this thesis, however, I will measure four emotions: anger, fear, sadness and joy. I will measure these four and no others, since, as outlined in Section 2.3.1, evidence suggests that these are the most basic, cross-culturally stable, emotions detectable in humans or even in other animals (see (Gu et al. 2019)). Following this reasoning, I will not attempt to measure "contempt" in this work, mainly for two reasons: 1) Under the taxonomy used here, contempt is a "complex emotion", composed of ideations, memories and other simpler emotions; and 2) Contempt is harder to detect in text data than simpler emotions, mainly due to its complexity. Complex emotions require more context to be interpreted, and they are hard to infer in text even for humans. Machine learning based emotion-detection techniques will have much more trouble detecting contempt than detecting simpler, more basic emotions, which ultimately will affect the accuracy of the predictions.

- **Group-based motivators**: The notion of a group-based or group-level motivator is

present in all of the reviewed socio-psychological models of collective actions. This group-based motivator is generally described as some kind of adverse situation inflicted upon the group the participant belongs to. In the SIMCA model (see Section 2.4.1) the group-based motivator was described as an "Injustice" perceived by the participants. Drury and Reicher (1999) and Klandermans and Stekelenburg (2013) defined "group grievances" as the group-based motivators, while Van Zomeren, Leach, et al. (2012) call them "collective disadvantages." What is clear and common to all models is that there is a group-level adversity in place. This adversity triggers the next step in the motivational process: an emotional reaction.

- **Collective Actions**: The collective actions performed by each of the social movements are a central factor in all of the models reviewed. However, only one of the reviewed models (see Tausch, Becker, Spears, et al. (2011b)) differentiate between types of collective actions and the motivational pathways by which individuals feel motivated to follow one or the other. In this work I will take the idea of differentiating between types of collective actions, but I will depart from the concept of "normativity" used by Tausch, Becker, Spears, et al. (2011b), since, as I outline in previously (see Section 2.2.2, I consider the axes of violent vs non-violent, and high-cost vs low-cost, to be clearer for the purposes of this thesis. High-cost collective actions usually involve the physical involvement of the individuals. Examples of them are: demonstrations, sit-ins and strikes in the non-violent side; and riots, arson and barricades in the violent side. High-cost collective actions can be considered "high-profile" as they are likely to be reported by the news media. Low-cost collective actions, on the other hand, usually don't involve the physical presence of the individuals. Examples on the non-violent side are: complaint letters, online campaigns and petitions. Examples of violent low-cost collective actions are: the spread of ill intentioned rumours or misinformation, collective attack of online services, and concerted online verbal attacks over individuals. Low-cost collective actions can be considered "low-profile", as many of them are not reported by the general news media. Consequently, in this work I will focus only on high-cost violent and non-violent collective actions, and how different emotions could differentially predict one or the other.

- **Ingroup**: Although not explicitly mentioned in every socio-psychological model of collective action reviewed above, the presence of a group identity defining an *ingroup* is essential

in the conceptualisation of social movements under the umbrella of the Social Identity Theory (see Section 2.3.4). Without an ingroup, group-level grievances are meaningless.

- **Outgroup**: With the definition of an *ingroup*, the instantaneous creation of the *outgroup* is a logical necessity, as it represents individuals not in the ingroup. However, here a more narrow concept of outgroup will be applied, consisting of adversaries to the social movement, following the Integrative Model of Collective Actions by Becker and Tausch (2015).

It is important to keep in mind that the socio-psychological models I reviewed in this chapter refer to individual's behaviours and individual's cognitive processes. E.g., Social Identity Theory, although a theory trying to explain group behaviour, does so by analysing how individual characteristics explain the emergence of group behaviour. In this work I am attempting to extend the findings and conclusions of individual level models to aggregated data. This is not without precedent, in fact some of the theories reviewed already did that. For example, the Inter-Group Emotion Theory (Mackie et al. 2008) suggests that while emotions are felt by every individual, their aggregation could be interpreted at group-level emotions. Nevertheless, it is worth acknowledging that there is a gap between various levels of analysis, i.e., the application of individual level theories to aggregate level data, which can be problematic. A common example of this is the question of "emergent behaviour" of aggregated systems, based on the simple rules of its components (Smith and Conrey 2007).

Additionally, in order to move forward I need to make some assumptions that will help the reader to understand better the research conducted in this thesis. I'm going to assume that the aspects described above are mental representations that "live" inside the minds of the participants of the social movements. Each of them have mental representations of what is its ingroup, outgroup, group-based motivation, and the collective actions they carried out. I will assume that each of them can feel and express one of the four basic emotions (anger, fear, sadness, joy) towards those aspects. I will also assume that such emotional expressions are, in part, reflected on their tweets, and that they can be detected. However, my main dependent variable is none of them, but rather the reports done by third parties (in this case, news media agencies) about the collective actions carried out by such social movements. To rephrase, I'm trying to predict real world events of collective actions, based on the emotions participants of social movements have over these four aspects (which also include the emotions participants

can have about the collective actions themselves).

The diagram below describes, hopefully in a clearer way, the model I'm set to test.

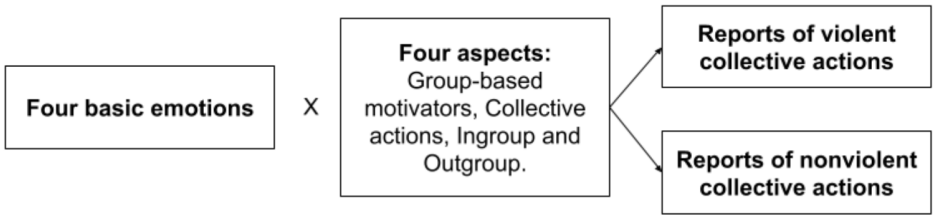

Figure 2.5: Model representing the analysis carried out in this thesis. Source: Personal collection.

Note that the above diagram is not a suggestion of a model to be tested statistically. Instead it tries to represent all the possible combinations of pairs of emotion-aspects that can predict the onset of violent or non-violent collective actions.

The full list of possible relations is quite extensive, whit a total of 32 possible hypothetical assumptions. Although all these assumptions will be empirically examined, it is important to note that this research is not aiming to test hypotheses. Computational approaches with observational data are typically data-driven and exploratory in nature (see Lazer et al. (2009) and Hofman et al. (2021) for a more integrative view). Nevertheless, a non-exhaustive set of general hypotheses based on the general predictions that can be derived from the socio-psychological models reviewed in this chapter will guide the research presented here:

1. Anger will generally be positively associated with non-violent collective actions. Following the evidence provided by the socio-psychological literature on collective actions, anger is seen as a motivator of collective actions (Tausch and Becker 2013). According to Tausch, Becker, Spears, et al. (2011b) anger is also an emotion aimed to heal broken relationships, leading generally to non-violent collective actions, as opposed to, for example, "contempt" (Tausch and Becker 2013) which is a more destructive emotion, leading generally to violent actions. Specifically:

    - Anger towards the outgroup will be positively associated with non-violent collective actions. I.e., the participants are angry with the outgroup but still look to repair the relationship between ingroup and outgroup.

- Anger towards the group-based motivator will be positively associated with non-violent collective actions. I.e., the participants are angry about the perceived illegitimate situation but still look to repair and modify the social conditions that lead to that grievance.
- Anger towards the ingroup will be negatively associated with violent collective actions. I.e., the participants are angry with either themselves or the general state of the movement, hence there is no stable, united foundation for risky violent actions.
- Anger towards the collective actions will be negatively associated with violent collective actions. I.e., the participants are angry about the actions taken by the movement and hence, participation in violent actions is discouraged.

2. Fear would generally be negatively associated with violent and non-violent collective actions. According to the literature reviewed in this chapter, fear is an inhibitory emotions that should decrease the motivation of individuals to participate in any action. Specifically:
   - Fear towards the collective actions will be negatively associated with any type of collective actions. I.e., the participants evaluate that the action taken by the movement are risky and hence participation is discouraged.
   - Fear towards the outgroup will be negatively associated with any type of collective actions. I.e., the participants fear the outgroup's reaction, and hence participation is discouraged.
3. Sadness is not explicitly mentioned in the socio-psychological model reviewed in Chapter 2, however, a group level grief stage is always mentioned prior to the onset of group-based anger. This grief can be related to any of the four aspects (ingroup, outgroup, group-based motivator and collective actions). However, existing literature does not suggest that the target of the sadness will change the outcome. The effect of sadness with respect to any of the four aspects will therefore be explored without any prior hypothetical assumptions.
4. Finally, joy, as self-directed positive emotion, has been hypothesised to have a positive effect on the participation in collective actions (Becker, Tausch, and Wagner 2011b). Based on this, I expect that general joy, joy towards the ingroup (self-directed) and joy towards the collective actions (self-directed) to positively predict participation in collective actions,

violent or non-violent.

Finally, the reader must have noticed that the four aspects have a peculiarity that emotions do not have. While the four basic emotions were chosen because of their universality across cultures, allowing them to be expressed in roughly the same way no matter the context, exactly the opposite is true for the four aspects, which are highly dependent on the context and on the social movement to which they are referring. Consequently, in the next chapter I will describe each social movement's context and circumstances and I will attempt to identify the four aspects for each of them.

Chapter 3

Tracked Social Movements

In this chapter I will characterise the tracked social movements chosen to be analysed in this project. The choice of social movements to be tracked was partially constrained by two main factors: the availability of Twitter activity of the social movement; and the availability of collective action reports for each of the social movements. These two factors are crucial, since without any of them, the relationships between the emotions expressed by social movements on social media cannot be related to the real world collective actions performed by them.

For each social movement, I will provide a characterisation of the four aspects outlined in Chapter 4, Section 2.5, of social movements that will become later the main targets of the emotional response analysis.

3.1 The Chilean 2019 Social Movement

The 18th October of 2019 marked the start of a widespread social unrest in Chile, unseen in decades. The people of the South American nation took their anger to the streets by the millions, staging massive protests and marches, several of which ended up in very disruptive riots and looting. They demanded greater social equality, improvement of the pension system, and a change in the constitution imposed during the dictatorship era. In order to better understand the origin and motivation of these social movements I will first review its group-based motivations.

3.1.1 Chilean 2019 Social Movement's Group-based Motivator

The main motivation for the social uprising is a combination of historical grievances of the working and middle class of Chile, that have dragged on for several decades; and contingent events occurring during the first half of the government of Sebastian Piñera in 2019.

On the side of historical grievances, Chile has had a complicated recent history when dealing with social equality demands. The most salient —and probably most famous— event is the military coup in 1973 against President Salvador Allende, led by Augusto Pinochet, who later became a dictator for 17 years. Pinochet's dictatorship scrapped many of the social reforms pushed by Allende aimed to increase equality, in favour of a very liberal economic model in which privatisation of basic services was the norm. This led to a great increase in inequality and segregation in the country during the dictatorship. This event still resonates within the working and lower middle class in Chile, and the figure of Salvador Allende is seen as a symbol of commitment to social equality, while Pinochet and his collaborators are seen as symbols of oppression and segregation. This was relevant for Piñera's 2019 government. Among his ministers were prominent promoters of Pinochet's social and economic reforms. His cabinet was perceived as more ideological than the one seen in his first government (Kozak 2018).

With the return to democracy in 1990, working class Chileans saw their hopes for greater social equality renewed. The "Concertacion" coalition, which won the 1989-1990 presidential elections, promised the return to the social-based policies under the slogan "Chile, la alegria ya viene" (Chile, the joy is coming). The Concertacion was able to remain in power for 20 years, from 1990 until the 2010, when the first mandate of Sebastian Piñera broke the winning streak of the centre-left coalition in the presidential elections. During those 20 years, Concertacion presidents focused their efforts on increasing the economic stability of Chile, while keeping a responsible expenditure on social programs (Fernandez and Vera 2012). Their policies led Chile to become one of the more stable and attractive countries for foreign investment in the region, and to join the OECD. In macroeconomic terms, Chile was seen as a prosperous and responsible country. However, several deep social problems remained unaddressed by the policies pushed by Concertacion, including increasing income inequality, a heavy increase in the cost of living and the unaddressed problem of segregated education, healthcare and pension systems (Fernandez and Vera 2012). These problems were partly addressed by improvements in credit access for the working and middle class. However, this was seen by the people as a way of transferring

the responsibility of the state to take care of the citizens, to the citizens themselves. These conditions created a highly indebted middle and working class, who felt that they were working "just to pay their debts." Although these problems were partially addressed by several bills protecting the rights of the customers and citizens against predatory loans, this was not enough to suppress the feeling of the working and middle class of being overwhelmed by debt.

While in his first mandate, Piñera followed basically the same model of tamed social expenditure used by the Concertacion, during his second mandate he was adamant about focusing on economic growth and security at the cost of social expenditure. With this agenda in mind he brought back several figures with ties to the former dictatorship, in a move that was heavily criticised by the public (Kozak 2018). In this environment of a stressed working and middle class, and of animosity towards government officials, the Transport Minister Gloria Hutt announced a raise of $30 pesos (£0.03) in metro fares for the whole network of the capital Santiago on the 1st of October of 2019. This seemingly low increase was received with dismay among the working class population, who regularly use the metro system to move around the great extension of the city of Santiago.

3.1.2 Chilean 2019 Social Movement's Collective Actions

In response to the Metro fare increase, on the 7th of October of 2019, secondary school students initiated a campaign of fare dodging under the war cry "¡Evade!". Piñera's government responded with disdain, arguing that students were not affected by the raise. An argument that was perceived as patronising by the students, whose parents were directly hit by the increase in fares. Since the students did not concede their protest, Sebastian Piñera, through Gloria Hutt (Transport Minister), instructed the riot police to protect the Metro stations. They were to ensure everyone paid the fare, and would threaten the students with applying the Internal Security Act, a law reserved only to penalise terrorist acts inside the country.

The riot police's forceful approach was not well received by the public, and the images and videos on social media of policemen mistreating secondary students caused widespread outrage in the middle and working class population. On the evening of the 18th of October of 2019 thousands turned to the streets to protest against the excessive repression against the students, and many took out their anger against the metro itself. Over 100 stations suffered near simultaneous attacks of diverse magnitude. Dozens were reduced to ashes.

During the course of the week from the 20th to 25th the demonstrations grew in numbers day after day. As a response, the repression also increased, leaving many citizens injured and blinded by the pellets shot by the riot police, and some killed by the action of the security forces. After weeks of protests during October and November of 2019, the proposal of a referendum to change the dictator's constitution was the gesture of goodwill necessary to calm down the demonstrators. Besides, the rewriting of the constitution was a symbolic goal that could be achieved in a relatively short period of time. However, this gesture only helped to reduce the size of the demonstrations, as regular Friday marches that in many cases turned into a confrontation with security forces, continued to take place for several months until the beginning of lockdown measures due to the Covid-19 pandemic in March 2020.

Figure 3.1: View of the demonstration in Santiago de Chile, October 2019. By Hugo Morales. CC BY-SA 4.0

3.1.3 Chilean 2019 Social Movement's Ingroup

The 2019 Chilean social movement was a decentralised social movement. There was no single political group nor individual who could have successfully claimed leaderships or full representatives of the demands of the movement. Moreover, as the social movement grew in size and magnitudes, several new demands emerged from various participating social movements, such as the Feminist Movement, the Indigenous Movement, and the Environmental Movement. This

makes it difficult to establish who is part of the ingroup of this greater social movement, since in this particular case the movement is more diffuse than other social movements of interest here. Still, some cautious descriptions can be made about the ingroup of this social movement:

- A leftist movement: The participants of the movement would not affiliate themselves with any current political party in Chile, but generally speaking the movement displays greater sympathies with left-wing parties and ideas. The figure of Salvador Allende was present in most of the demonstrations and it was common to see left-wing politicians and artists participating in the marches.
- A middle and working class movement: For readers not familiar with the wealth distribution in Chile, a middle and working class social movement, whose main concern is more social equality, might be confusing. However, in Chile the wealth distribution makes the distinction between working and middle class rather blurred. Chile remains one of the most unequal countries in the OECD, before and after tax redistribution. Most middle class people are professionals with tertiary educational degrees, but many of them are the first generation within their families, who had the opportunity to attend university. This tied them emotionally to relatives and friends, who are in a more precarious economic position.
- Students: Secondary school and tertiary education students actively participated in the social movement and were, especially at the beginning, the most iconic image of the protest. Secondary school students were the trigger of further mobilisations with their fare evasion campaign.

3.1.4 Chilean 2019 Social Movement's Outgroup

To distinguish the output of the 2019 Chilean social movement, under the conceptualisation made in 3, I will try to identify the main actors, who opposed the main demands of the movement. These actors could be authorities, individuals, institutions or groups of people.

- The government: The government was the main antagonist of the social movement. This was evident since the start of the mobilisations in which protesters demanded the resignation of ministers, police chiefs and the president himself. The status of the government as the main adversary to the social movement remained until March 2022, when Gabriel Boric was elected to be the new president of the nation. Gabriel Boric, was a former

student representative and activist during the student protest in the years 2011 to 2013. In 2013 he was elected as a member of the Chamber of Representatives as an independent candidate, and later in 2017 was reelected as a member of the "Broad Front" leftist coalition. Boric was largely in favour of most of the demands of the 2019 social movement, and thanks to this was seen as a conciliatory politician and a bridge between the demands of the protesters and the political class.

- Police and security forces: The police were seen as a major source of repression by the participants of the social movement. This has a historical component, since Carabineros de Chile (uniformed police of Chile) were actively involved in the military coup against the government of Salvador Allende, and later, as an institution, participated in violations of human rights against opponents of the dictatorship of Augusto Pinochet. Additionally, they acted negligently and violently when trying to dissolve demonstrations, using extreme force against protesters, many of whom were injured and permanently scarred.
- Right-wing opposition: Given that the movement identifies itself with left-wing ideologies and left-wing historical figures, right-wing individuals were swiftly targeted as adversaries of the social movement. And although this categorisation was initially not fully shared by right-wing individuals, they eventually identified themselves as an opposition to the social movement and its activism strategies. This was reflected in prominent conservative figures from politics and industry heavily criticising the violent character of some demonstrations and their impact on the economy of the country.
- Retail companies: Finally, several retail companies were seen as antagonists of the social movement due to their participation in recent high profile cases of corruption and price collusion. These cases, particularly the collusion in manipulating prices of food, basic household products and basic medications were bitterly resented by working and middle class Chileans. For this reason, many supermarkets and pharmacies belonging to these retail companies were the target of major looting and attacks in the first weeks of the social uprising.

3.2 The 2019 Hong Kong Social Movement

During most of the year 2019 and the beginning of 2020, Hong Kong witnessed a series of massive demonstrations and protests in response to the Fugitive Offenders and Mutual Legal Assistance in Criminal Matters Legislation (Amendment) Bill 2019. The bill was proposed by the Hong Kong government to provide mechanisms for the transfer and extradition of criminal offenders between Taiwan, Macau and mainland China.

The background for such a proposition goes back to 1997. In the final months of British rule, Hong Kong legislators passed laws preventing the extradition of Hong Kong citizens to mainland China. This was motivated by fears over losing freedom of speech and civil and human rights under the new "One Country, Two Systems" agreement between Hong Kong and mainland China. These laws were not well received by Beijing, which started planning to reverse the bills right after the handover, which would allow the Chinese government a streamlined way of prosecuting individuals in Hong Kong and to reduce the use of "extraordinary measures," which often included the kidnapping of Hong Kong citizens to moved them to mainland China territory and proceed to their official detention. (Lague et al. 2019).

After a high-profile murder case in Taipei (Taiwan) in early 2018, involving Hong Kong citizens, ended without criminal charges by authorities from either of countries involved (Siu 2019), Hong Kong and Taiwanese authorities were pushed by public opinion to device mechanisms to prevent crimes committed by Hong Kong citizens abroad to go unpunished. With this overarching goal in mind, the Fugitive Offenders and Mutual Legal Assistance in Criminal Matters Legislation Bill was proposed. If the bill was to be passed, Taiwan, Macao and, most notably, China would have a way to lawfully prosecute and arrest offenders on Hong Kong territory with the aid of Hong Kong authorities. This caused great concern among several Hong Kong groups and organisations, who saw this bill as a way for Beijing to submit Hong Kong citizens to the mainland legal system.

3.2.1 2019 Hong Kong Protest Group-based Motivator

As mentioned above, the immediate cause for the protest events in Hong Kong during 2019 and 2020 was the proposition of the extradition bill with Taiwan, Macau and China. Especially of concern was establishing a mechanism of extradition with mainland China, as this was seen by Hong Kong citizens as an advance of Beijing in establishing their legal systems in the city,

which would possibly result in the loss of individual rights and freedoms (Chernin 2019).

In addition to the extradition bill, a growing sentiment of distrust towards mainland China had brewed in Hong Kong since 2014 after the so-called "Umbrella Revolution" (Cheung 2019). The Umbrella Revolution, also called the 2014 Hong Kong Protests, was a series of demonstrations and sit-ins in the city of Hong Kong between the 26th of September and the 15th of December of 2014. The protest began as a response to a proposed reform of the Hong Kong electoral system, issued by the Standing Committee of the National People's Congress. This reform was seen as complacent with the intention of the Chinese Communist party to screen candidates previous of the election. The demands of the Umbrella Revolution were: Genuine universal suffrage; Retraction of the Standing Committee of the National People's Congress issued reform; Abolition of functional constituencies [1] of Legislative Council of Hong Kong; and Resignation of the then Chief Executive of Hong Kong Leung Chun-ying. None of the demands yielded any concession. The failure of the Umbrella Revolution in bringing concessions left young activists uncertain about the level of willingness of the Chinese authorities to ever negotiate any concession with the people of Hong Kong. And with the special status of the city expiring in the coming decades, many feared for the future of the city (Cheung 2019). This fear was not entirely unjustified. After 2014 (Leung 2019) Chinese authorities increased their efforts on silencing dissident voices in Hong Kong, issuing unlawful arrests and criminally prosecuting activists (Cheng 2016).

All the above mentioned concerns where then crystallised in the following five demands of the 2019 Hong Kong Protest Movement (Kobayashi et al. 2021):

- Withdrawal of the extradition bill from the legislative process.
- Stop the characterisation of the protests as "riots" by government officials. The characterisation of riots brought considerably higher penalties.
- Amnesty for arrested protesters. Some protesters were arrested in hospitals, violating the confidentiality of medical records.
- Establishment of an independent Commission of Inquiry into alleged police brutality.
- Resignation of Chief Executive of Hong Kong Carrie Lam, and the implementation of

[1] In the political system of Hong Kong, a functional constituency is a professional or special interest group involved in the electoral process. Eligible voters in a functional constituency could include ordinary citizens as well as other entities such as organisations and corporations.

universal suffrage for the Legislative Council [2] elections and for the election of the Chief Executive.

In this work I will consider these demands and the references to them as the main group-based motivators of the Hong Kong movement.

3.2.2 2019 Hong Kong Protest Collective Actions

While the initial protests and demonstrations started in March - April 2019; 9th of June of 2019 was the day when the first large scale demonstration took place. Organisers estimated 1.03 million people attended, while the police only estimated 270,000. Even considering the drastically lower estimation of the Hong Kong police, a demonstration including hundreds of thousands of participants was something not seen in decades in the city (Lague et al. 2019). In addition to this, on the 12th of June a second demonstration, gathering around 40,000 protesters took place at the Legislative Council of Hong Kong (LegCo) complex. These two demonstrations successfully stopped the discussion of the extradition bill at LegCo. However, police authorities characterised the event as a riot, which resulted in heavy repression against the protesters and higher penalties for the arrested activists. The policy repression against the protesters sparked heavy criticisms from the public, which culminated in the demonstration on 16th of June, which gathered around 2 million people at Victoria Park, Causeway Bay, according to the organisers (Reporters 2019). After this march, the demonstration spread to other neighbourhoods of Hong Kong.

On 21st of July 2019, a group of unidentified armed individuals, in white clothes, assaulted people indiscriminately inside the Yuen Long metro station. The attackers were clearly sympathisers with mainland China, and their main target were attendants of a pro-Hong-Kong demonstration (Standard 2019). Despite thousands of calls, police took around 40 minutes to arrive at the scene, and no white-dressed attackers were arrested. This incident marked a turning point in the protest. The confidence in the police dropped as they were seen as collaborating with the attackers and in clear support of the pro-China camp. This caused many of the more apathetic citizens to take side of the Hong Kong movement (Purbrick 2019), increasing further

[2] The Legislative Council of the Hong Kong Special Administrative Region (LegCo), is the unicameral legislature body of Hong Kong, who sits under the "One country, two systems" agreement between Hong Kong and mainland China. Among its powers, the Legislative Council have the ability to enact, amend or repeal laws; approve budget and public expenditure; and endorse the appointment or removal of the judges of the Court of Final Appeal and the Chief Judge of the High Court, as well as the power to impeach the Chief Executive of Hong Kong

the animosity towards the security forces. This became evident on October 1st, when several major massive protests and violent clashes with police took place during the 70th anniversary of the foundation of the People's Republic of China. This was the first time the Hong Kong police used live rounds against protester (Lam et al. 2019). Similar events continued to happen during October and November 2019.

Figure 3.2: Hong Kong demonstration, June 2019. By Wongan4614. CC BY-SA 4.0

The 2019 Hong Kong movement was characterised as having highly organised tactics and methods, despite its leaderless nature. Their method of organisation included the use of anonymous forums and social media platforms (e.g., Reddit), and the use of end-to-end messaging applications (e.g., Telegram), in order to avoid identification by authorities. In general terms, two main groups can be identified:

- The moderate group: comprising the majority of the protesters, the moderate group tactics and methods were mostly peaceful and non-disruptive. Common collective actions included mass rallies, petitions, hunger strikes, labour strikes, artistic expressions, calls for international support, among others.
- The "fighters:" Radical protesters, known as 'the fighters' (Kuo 2019), adopted a 'be water' strategy, taking inspiration from the famous Hong Kong movie star Bruce Lee. The idea of the 'be water' strategy was to avoid direct confrontation with the police when possible in order to minimise the risk of arrest. They achieved this by retreating when

> police arrived and then reemerging in other locations of the city. There was considerable coordination among the "fighters," with some individuals in charge of the front line and general protection of the protesters, others in charge of first aid, and again others in charge of the supply chain of water bottles, umbrellas and other articles. The 'fighters' group collective actions were considerably more disruptive and violent. In cases in which confrontation with police was the goal, protesters threw rocks, petrol bombs and corrosive agents at police forces, as well as engaging in targeted looting and arson attacks on stores and business offices known to support mainland China.

Although being groups with drastically different methods, there were no reports of major conflicts between the "fighters" and the moderate group.

3.2.3 2019 Hong Kong Protest Ingroup

The Hong Kong protest movement, in contrast with the other movements studied in this project, does have a more clear differentiation between the ingroup and the outgroup. The fact that the conflict itself has a geographical component does contribute to the differentiation. However, the geographical component is not the whole picture. The Hong Kong protest movement has positioned itself in opposition to mainland China policies. Pro-democracy and pro civil and human rights organisations are among the most prominent actors inside the movement and they are in clear opposition to the policies promoted by Beijing (Purbrick 2019; Holbig 2020). There is also the sense that Hong Kong has lost its status of a world capital. The decreased number of Western residents and simultaneously the increased number of Chinese mainlanders has fuelled the sentiment of Chinese re-colonisation among Hong Kong citizens (Holbig 2020). The increased influx of Chinese mainlanders has also affected the real-estate and housing market, already very stressed in Honk Kong. This has affected especially young Hong Kongers who have seen prices of housing increase dramatically in recent years (Holbig 2020).

Considering the above mentioned factors, the 2019 Hong Kong Protest ingroup can be described as mainly citizens of Hong Kong (with the occasional supporter coming from other cities or countries), living in Hong Kong or in the surrounding territories. They are also mainly young (under 30) and/or students, holding pro-democratic and pro civil rights values, and with generally pro-western inclinations.

3.2.4 2019 Hong Kong Protest Outgroup

The parties opposing the 2019 protest in Hong Kong were multiple, although tied together by their relationship with mainland China. The front line of the adversaries consisted of the local government and authorities of Hong Kong, which included: the Hong Kong Government Secretariat, the Hong Kong Executive Council, the Hong Kong Police Force.

There were also some extra-governmental organisations supporting the local authorities (e.g., Safeguard Hong Kong Alliance, Politik Social Strategic), but the main support came from the Government of China and the People's Liberation Army (Hong Kong garrison). All these actors were seen as the main opponents to the demands of the Hong Kong protesters.

3.3 Fridays For Future

Fridays For Future (FFF), or School Strike for Climate (Swedish original: Skolstrejk för klimatet), is an international social movement of school-age students who skip Friday classes to protest against the inaction of authorities against climate change. The movement started with young Swedish activist Greta Thunberg who, while being in ninth grade, staged a protest outside the Swedish Parliament with a sign reading "Skolstrejk för klimatet" in August 2018. Her demands were for the Swedish government to reduce the country's emissions in order to align with the Paris Agreement. In 2018 she was invited to speak at the plenary session at COP24 in Katowice Poland, which helped to establish her as an international personality. In September 2019, Greta travelled to New York city in a two week journey by sail boat to speak at the UN Climate Summit, the 23rd of September of 2019. After this intervention, Thunberg gained increased international notoriety, inspiring students in several countries to start their own strikes for climate.

The movement itself is inherently decentralised. While Greta Thunberg served as the inspiration for many of the members, she does not lead the movement and several other leading young activists are taking active participation in the movement.

3.3.1 Fridays For Future Group-based Motivator

The main motivator of the FFF is the current man-made climate emergency (IPCC 2022; IPCC 2014). The analysis made by Spaiser, Nisbett, et al. (2022) makes this clear, adding that

the FFF was able to establish themselves, the younger generations, as victims of the climate emergency. Many if not the majority of the participants of FFF do feel and believe that their future livelihood is at stake because of governments' inaction against climate change (De Moor et al. 2020; Spaiser, Nisbett, et al. 2022), which makes their motives much more concrete and pressing. Additionally, Spaiser, Nisbett, et al. (2022) points out that for FFF "allowing the clime crisis to unfold is deliberately facilitating atrocities (e.g., genocide by famine)." This suggests that avoiding atrocities and mass human rights violations caused by the climate emergency is also one the group-based motivators of FFF.

With all the motivations of FFF translate into three general demands, as stated on their website (FFF 2022), are:

- Keep the global temperature rise below 1.5 °C compared to pre-industrial levels.
- Ensure climate justice and equity.
- Listen to the best united science currently available.

Notably and since the beginning, FFF included in its demands the need for climate justice and support for climate refugees.

3.3.2 Fridays For Future Collective Actions

Although not self-defined as such, Fridays For Future is undoubtedly a non-violent social movement (De Moor et al. 2020). Its actions are carried out by students of middle to high-school age, with the participation of university students in some countries, who have little intention to act violently against police or other authorities. Their actions include mainly sit-ins at administrative centres and offices in the different countries and cities where they are carried out, and/or class-skipping in order to go on marches and demonstrations, i.e., school strikes. This strategy has been widely successful, allowing the movement to have participants in many countries, and notably, allowing the movement to have presence in the Global South, having a significant amount of actions and participants from Africa, South Asia and South-East Asia. The most important massive actions carried out so far by the movement were performed in September 2019 during the Global Week for Future and in November 2019 before the beginning of the COP25 in Madrid. According to organisers, the global strike on the 20th of September 2019 involved more than 4 million participants in over 4500 locations across 150 countries, while the

global strike on the 27th of the same month reported over 2 millions participants. The global strike carried out on the 29th of November 2019, three days before the start of the COP25, reported more than 2 million participants, in 2400 cities across 157 countries.

Figure 3.3: Fridays For Future school strike in Toronto, March 2019. By Dina Dong. CC BY-SA 4.0

3.3.3 Fridays For Future Ingroup

The main ingroup of Fridays For Future are school/University students, who are taking part in the FFF protests. The movement defines itself as a student movement and its actions are tailored to be carried out by (mainly) secondary school students. Other activist groups (e.g., Extinction Rebellion, Sunrise Movement) can be considered allies of the movements. Overall, there is no doubt that FFF is a youth-led movement (De Moor et al. 2020). Additionally, De Moor et al. (2020) shows that the majority of the FFF participants are identify themselves as female (69%).

In addition to the young, school age students, there is also a good proportion of adults who identify themselves as part of the movement. De Moor et al. (2020) shows in their study that up to 69% of the participants of FFF are above 19 years old, and this proportion is thought to keep growing as the movements manage to convince older generations to join.

Finally, De Moor et al. (2020) shows that participants above the age 26 have a high education level, with around 70% of them having some kind of university degree.

There is no political affiliation of the movement, but there is an explicit awareness of the necessity of climate justice and support for climate refugees, which can be considered left-leaning or progressive.

3.3.4 Fridays For Future Outgroup

The main outgroup for Fridays for Future are governments and authorities as the main demands are being directed at them. Governments and authorities are also seen as enablers of the climate emergency because of their lack of action (Spaiser, Nisbett, et al. 2022; De Moor et al. 2020). Fossil fuel companies are a secondary outgroup of the movement, since although they are seen as the main cause of the climate emergency, FFF do not expect anything from them (Spaiser, Nisbett, et al. 2022). Consequently, their communication and collective action are intended to sway public opinion to push governments to take action against the fossil fuel companies and other industries related to global warming and climate change emissions. Their main form of activism, the school strike, has also been the source of criticisms by right-wing politicians and press (Nevett 2019) adding them to the list of antagonists of Fridays For Future. Finally, older generations (e.g., Baby Boomers) have been pointed by the movement as enablers of the climate emergency and, in some sense, responsible for the "destruction of their future."

Chapter 4

Methodological Framework

This chapter presents the procedures of data collection, data transformation, data storage and choice of analytical techniques used in this dissertation.

The use of social media data and computational methods to answer questions related to social phenomena has seen an increase in the last decade leading to the advent of the sub-field of Computational Social Science, an interdisciplinary discipline combining the technical abilities of computer scientists, with the interest in social science research questions. However, and in part due to the nature in which the data are produced, computer scientists have taken the lead in the sub-field of computational social science (Bonenfant and Meurs 2020). Theocharis and Jungherr (2021), propose this limitation should be seen as an incentive for interdisciplinarity, allowing the creation of a common set of standards for the discipline without the need to retrain social scientists into "mediocre coders" or turning computer scientist into "mediocre social scientist", which is seen as the most viable option to cope with the challenges this new field poses. While I agree with this proposal, I disagree with the assumption that social scientists should leave the technicalities of the discipline in the hands of the "technicians". The technical opaqueness of the field of AI and Machine Learning has created many problems in this field in recent years, and Computational Social Science should not make the mistake of ignoring technicalities which might have important ethical and political consequences.

Social scientists venturing into the realms of social media data and computational methods should follow the example of other disciplines that faced similar problems in the past. For example, astrophysicists, in order to make sense of the large quantities of data provided by

their measurement instruments, had to learn not only how to use computers to aid them, but also how to code and how to deal with the different types of data. We should imitate them and, in an incremental manner, we should also fully understand how to use the available tools to make social research and how those tools also can limit what we can do. It is the only way in which we can be fully responsible for the conclusion we can draw from using computational methods.

4.1 Data sources and data collection procedures

Since one of the main objectives of this project is to capture the emotional variation over time of different social movements, social media posts (data) related to said movements is a good proxy to track their evolution. However, not every social media site was equally ideal for this task. The platform had to be popular enough to capture an important user base participating in the social movements. The platform should also allow users to share an analysable type of content in order to have a greater chance to capture emotional reactions from them. Platforms focusing on sharing images (e.g., Instagram), in this case, were quickly discarded, as it is considerably more difficult to infer emotional reactions from shared images alone. Social media platforms allowing their users to share a variety of content, but also to react to it in the form of text comments, are ideal, as they provide text data produced by users, which is easier to collect, store and analyse.

Another important aspect to consider when choosing social media platforms from which to collect data is their Terms of Service (ToS). ToS are a simple (yet in some cases very exhaustive) legal agreement between the service provider and the users. In this case, social media platforms are the service provider. They provide the platform in which users can interact (e.g., share images, comments, or videos; make friends; re-share and like other user's content). But they also provided utilities and services for secondary user bases who might want to advertise on the platform or collect data from it. This means that if a third party wishes to collect data from a social media platform, all the data collected and the methods to be used in the data collection process must be approved by the ToS of the platform. Otherwise, the data collection process might be in violation of the user rights and the data could be considered invalid for research purposes. ToS are also regularly updated, adding or removing restrictions in which the platform and its data can be used. A good example of this is Facebook and its Graph API (Ap-

plication Programming Interface). Before 2016, Facebook provided developers, advertisers and researchers a method to collect data from its platform, including large amounts of information about its users, called "Graph API". This API provided a way to collect data from Facebook, which was compliant with their, at that point, very relaxed ToS. In that sense, Facebook would have been a great source of data for this project. Sadly, after the Cambridge Analytica scandal in 2016 (Isaak and Hanna 2018), Facebook changed its ToS, closed its API, and no longer allowed data collection from the platform for research purposes. And while it is possible to parse the data published on the site using web scrapers (programs capable of automatically collect data from what we see in the web browsers), that is not a legal use of the platform since it violates Facebook's ToS and the trust of the users, and depending of the country, might be punishable by law.

Finally, it has been shown that social media platforms demographics are not representative of the population (Mellon and Prosser 2017). In general, the user base of social media is younger and tends to be more educated than the general population and more willing to support left leaning politics (Mellon and Prosser 2017), while at the same time participating less in institutional ways of political participation (e.g., voting). Despite this lack of representativeness, the user base of social media platforms is big enough to be of interest for social research. According to Kemp (2021), in April 2021 the adoption of social media accelerated again, with an increase of 14% comparing to the same month in 2020, reaching a total number of active social media users of 4.33 billions, 55.1% of the Global population, with an average time spent on social media of 2 hours 22 minutes (Kemp 2021). In terms of the favourite social media platform, Kemp (2021), reports Facebook and associated services WhatsApp and Instagram, to be in the first, second and third positions respectively, followed by Twitter in fourth place. Popularity and a diverse user base is a very important factor when considering a social media platform as a source of data. A popular enough social media platform increases the probability to find activism related to the social movement being tracked in this project. A diverse social media platform increases the probability of finding diverse reactions to the activism being promoted by said social movements.

4.1.1 Twitter as a source of data

According to Kemp (2021), Twitter is the 4th favourite social media platform (without considering the Chinese market). In terms of user base, Twitter falls to position 15th, with an estimated

user base of 396 Millions. This is way below the 2.8 billion users of Facebook or the 2 billion users of WhatsApp and even below Reddit with 430 million users. However, as it was mentioned above, popularity is only one of the factors to consider. In terms of diversity, Facebook and WhatsApp are heavily skewed with respect to middle age and older generations. Instagram is skewed towards younger women (Kemp 2021), while Reddit is skewed heavily towards younger males (Sattelberg 2021). On the other hand, and according to Kemp (2021), the Twitter user base is skewed less towards younger generations but has a slightly higher male participation than female participation.

Finally, legal and safe access to the data produced by the social media platform is crucial. In this aspect, Twitter has the upper hand. Twitter provides access to its data via its API, which is compliant with its Terms of Service. This API was recently overhauled to include exclusive access to academics using Twitter data to do research with no associated cost (Twitter 2021c). In contrast, Facebook, and its associated platforms (Instagram, WhatsApp, Messenger), now under the "Meta" umbrella, closed their free and open APIs right after the scandal of Cambridge Analytica in 2016, and have only reopen commercial APIs (e.g., Meta Crowdtangle API). This limits the access to their data but also transparency of the research that can be done with their data. And while Reddit also has a very open and free to use API, their skewed user base and focus on technological topics, makes it an inferior option to Twitter in terms of source of data in the context of the research questions of interest here. This is shown in the consolidation of this platform and as a source of data for several research fields (Karami et al. 2020).

Collected Twitter Data

As specified in Section 4.1 tweets related to each social movement were collected using a series of keywords and hashtags (see appendix A) totalling around 500 millions tweets. After an initial process of data cleaning, the total number of tweets was reduced considerably to a total of 57.1 millions. This dramatic reduction is mainly explained by the exclusion of data related to tracked social events that never evolved to social movements (e.g., migrant caravans in the US, opposition to Jair Bolsonaro in Brazil), and the exclusion of social movements for which I could not find, despite my best efforts, a reliable source of real-world events data, as was the case with the Yellow Vest Movement ("Mouvement des gilets jaunes" in French) and Extinction Rebellion. For both of these cases the most complete list of events was found in their respective Wikipedia articles, which was still very incomplete and subject to the bias of the creator of the

articles. The final number of tweets per social movement included in the analysis can be seen in Table 4.1.

Social Movement	Number of Tweets
Chilean 2019 Social Movement	32,871,065
Honk Kong 2019 Social Movement	21,859,119
Fridays For Future	2,368,816

Table 4.1: Number of collected tweets per social movement

Prepossessing of Twitter Data

The collected Twitter data of each movement (see Table 4.1) was subjected to minimal pre-processing before being passed to the classification model for the detection of its underlying emotions. Unlike in the case of training data, in the main Twitter data, user tags and URLs were kept as well as all special characters and emojis. Tweets containing less than 5 words (including hashtags) were removed from the analysis to allow for a minimal amount of semantic content. Automatic padding (the addition of white-spaces to fill-up a predetermined sequence length) is the only pre-processing performed before the data were classified. This decision was made in order to preserve the maximum amount of information in the text.

The process of emotion classification was the first to be applied to the Twitter data of the three social movements. The process yielded a vector (a list of numbers) containing the probabilities of a tweet to express any of the four emotions. The probabilities were calculated using a *softmax* function. A softmax function σ (Bridle 1990) converts a vector $\mathbf{z}$ of K numbers into a probability distribution for K possible outcomes as follows:

$$\sigma(\mathbf{z})_i = \frac{\exp z_i}{\sum_{j=1}^{K} \exp z_j} \text{ for } i = 1 \ldots, K$$

The function takes every number z_i in the vector $\mathbf{z}$ and passes it to an exponential function, ensuring that the output is always positive, to divide every z_i by the sum of all the exponentials in $\mathbf{z}$. The output is constrained to a range between 0 and 1, so the sum of all the values in the output vector adds up to 1, allowing the interpretation of the values as probabilities. In our case, this means that if a tweet has a high probability of belonging to an emotion, it will necessarily have a smaller probability of belonging to any of the others. On the other hand, a totally ambiguous tweet will have a vector containing a very similar probability for each of the

emotions. Since here we are trying to detect four emotions, a totally ambiguous tweet should have a probability of 0.25 of belonging to each of the four emotions.

The behaviour of the softmax function used in the classification process allowed me to use the output probabilities as a threshold. In this thesis, only the tweets which achieved a probability above 0.8 in any of the emotions were considered for further analysis. Below this threshold the tweets were considered to be emotionally ambiguous. This reduced the amount of tweets to be used in the final analysis. The detail of this reduction is shown in Table 4.2 below.

Social Movement	Original N	N After filtering	% of Original Dataset
Chilean 2019	32871065	22500611	68%
Hong Kong 2019	21859119	14677228	67%
Fridays For Future	2368816	1620810	68%

Table 4.2: Number of tweets per social movement after filtering

These new datasets, containing the tweets and their respective emotion classification outcome, will be processed to detect terms related to the aspects of Group-based Motivator (GBM), Collective Actions, Ingroup or Outgroup, outlined and explained in Chapter 2. The detection will be performed using the set of keywords specifically selected for each social movement (see Section 5.3). If any of the words referring to one of the aspects (e.g., collective actions) appears in the text of a tweet, that word will be added to a counter measuring the frequency of that specific aspect in the tweet. The tweet will be considered to speak about a specific aspect by choosing the biggest absolute value among all the four counters. For example, if a tweet had the following counts per aspects: Group-based Motivator = 4, Collective Actions = 2, Ingroup = 2, Outgroup = 0; that tweet was considered to be speaking about Group-based Motivators. In the case of ties (two or more aspects have the same count) the tweet was considered *Ambiguous*. The cases in which the tweets had no mentions of any of the aspects were classified as *None*. This method implies that a tweet could contain valid mentions of more than one of the four aspects (Group-based motivator, Collective Actions, Ingroup and Outgroup), but still ensures that the selected aspect is the main topic of the tweet, by excluding tweets with ties in the counts. It also ensures that there are no duplicated counts between the aspects, so each selected tweet belongs to only one aspect. This strict selection methodology is necessary to eliminate ambiguities in the emotions detection process, in order to have a clear indicator of the emotions directed to each of the aspects.

An issue that needs to be addressed with respect to the data collection process is that there is no guarantee that everyone posting about the social movements is actually a participant in the social movement, they can be simply "bystanders" in the discussion. Under the conditions of the data collection process of this study, in which we do not have direct information about the protest participation of Twitter users, there is no exact mechanism that would allow to identify and discard bystanders from the sample. However, there are mechanisms that can help to clean the data as much possible. Here I used three methods:

1. Hashtag informed data collection: Hashtags are keywords of the format "#word" used by social media users to signal their post belongs to a certain topic. In Twitter in particular, hashtags can be used by users to search topics within the broader conversation and to signal communication campaigns within the platform. Social movements participants are well aware of this and they craft their own hashtags to signal their participation in the social movement. With sufficient knowledge of the social movement, one can use these hashtags to select tweets that talk about the social movements coming only form users signalling themselves as participants of the movement. This is not without problems. Hashtags can be hijacked by others groups for various reasons, so careful monitoring of the hashtags dynamics was necessary during data collection in order to stop data collection including hijacked hashtags.

2. Accounts informed data collection: Another method used in this study was to collect data from specific accounts, which were selected for their relation to the social movements. For example, in the case of Fridays for Future, besides the official Twitter account of the social movement, I included accounts declaring themselves as local chapters of the movement. As a precaution, I only included accounts that were followed by the main social movements account, which gave them some level of legitimacy.

3. "On-topic" pruning: When performing the data analysis by aspect, ambiguous tweets, and tweets that were not talking about any of the aspect of the social movements were dropped. This strategy helps to reduce tweets that are "off-topic" but are using the hashtag to gain visibility.

4.1.2 Events Data

As stated in Chapter 1, Twitter data from the tracked social movements has to be contrasted with the real life protest events in order to determine if there is any relation between collective action type and the emotions expressed by social movements on social media. For that purpose, events data are required, ideally a detailed record of all collective action events involving the respective social movements with information on the level of violence and magnitude of protest. Thankfully, this arduous task was already carried out by the The Armed Conflict Location & Event Data Project (ACLED)[1], which collects and curates news reports of contentious events occurring all over the world, providing date of occurrence, source, parties involved, and type for each event. ACLED exposes these data by using a data export tool, allowing researchers to download data in a tabular format. From the downloaded data, I considered the amount of news reports per day in the dataset as a proxy measure of the magnitude of the events. This database provided me with the events data for all the three movements in this study.

All events were binary coded as either violent and non-violent, based on the description provided by ACLED. Table 4.3 provides the number of news reports by social movement collected from ACLED.

Social Movement	Time frame	Violent	Non-violent
Chilean 2019 Social Movement	October 2019 to March 2020	731	870
Honk Kong 2019 Social Movement	August 2019 to March 2020	249	576
Fridays For Future	October 2018 to January 2020	0	342

Table 4.3: Number of violent and non-violent events **reported** for each social movement.

Based on Table 4.3 it can be seen that, the Hong Kong and Chilean 2019 Social Movements are the most similar in terms of news reports, with the Chilean 2019 showing a slightly higher proportion of violent events reports than the Hong Kong Movement. Finally, Fridays for Future, following their non-violence stance, only have non-violent reported events.

Although ACLED is a significantly better source of events data than, for example, Wikipedia articles, it does have some limitations. The main is that it relies on news reports to compile the information, leaving room for biases the news agencies can have over what is considered newsworthy. This might affect disproportionately non-violent collective actions, as they tend to be under-reported by news agencies. Sadly, I do not have any reliable way to compensate

[1] ACLED URL: https://acleddata.com

for this under-reporting, and thus the prediction made using these data should be taken as prediction over reported events, and not over the actual occurrence of events.

4.1.3 Data collection strategy

There are several ways in which data published on a website can be collected (Saurkar et al. 2018). Twitter provides an API to control the communication between the platform and third party applications collecting data from the site. Additionally, the data needs to be collected in a secure and reliable way. Finally, the data needs to be "ingested" before it can be used in analysis. The procedures to connect to the Twitter API, collect and store the data and finally ingest it are discussed in the following sections.

The Twitter API

Application Programming Interfaces are, as their name says, interfaces, but intended for communications between computer programs. As a general rule, an API allows communication between two computer applications. This communication often takes the forms of an exchange of services between a server application and a client application through the use of a purposely built communication protocol (Reddy 2011). These services can range from accessing stored data to interaction with physical devices. A very common example is the way web browsers (Chrome, Firefox, Safari, Edge) communicate with the physical devices of a computer. The communication is not direct. The web browser "asks" the operating system, through the use of an API, to get access to the network device, screen, audio, etc. In this way, the owner of the service, the operating system, can control the access to the device to prevent misuse by malicious programs. Understanding what an API is, helps to understand how APIs on social media sites work, since they are an extrapolation of these principles. They are designed to allow the creation of third party, general purpose, applications and hence the data are provided in a machine friendly format (JSON or XML). They usually require some sort of authentication as a way of controlling the access to the services provided, which are also usually separated between different "end-points" (addresses inside the API which help organise the different services provided by it).

Within Twitter's API, I opted to use the sampled streaming service end-point (Twitter 2021a). Twitter claims this service provides roughly a 1% random sample of all the tweets posted on the platform in real time. This stream was filtered using a series of keywords and hashtags (e.g.,

#ClimateEmergency, #hkprotest) (see appendix A) in order to obtain only the tweets related to the social movement being tracked in this project. While this procedure can be done in any programming language, in this case Python 3.6, and Tweepy 3.6.0 to 3.8.0 Python library was used (all code available in appendix A). The data collection procedure was carried out by a small Python subroutine, collecting real time tweets related to the three social movements over a period of approximately one year and six months, from November 2018 to February 2020, totalling approximately 500 million tweets.

Data storage pipeline

In order to safeguard the operation of data collection and the subsequent collected data, and in order to comply with the GDPR regulations, I decided to use an external, cloud based, data collection and storage solution inside a virtual machine (effectively a virtual computer assembled using the resources from a cluster of servers) in Google Cloud Services. Cloud Computing services provide several alternatives to automate the process of data collection and data storage, but since current privacy regulation require that the devices must be inside the jurisdiction of the United Kingdom, several options, namely distributed computing (e.g., Google BigTable), are not eligible, as there is no guarantee that the hardware is within the national territory.

But, complying with GDPR regulation, the physical location of the virtual machine was inside the UK. I opted to use a document-based database system (MongoDB) since this allows for the storage of unstructured data. Traditional, structured database solutions, like SQL-based systems, require to know the structure (the tables, fields, columns or variables) the user needs to store in advance, in what is known as a "database schema". While this is useful to prevent the insertion of unexpected data into a database, when dealing with JSON objects being sent from external APIs, which might suffer unexpected changes in structures, these solutions are sub-optimal. In contrast, document-based database systems accept arbitrary documents into collections. These documents can be of any structure and type. Although this introduces a potential for incomplete or damaged entries, for data collection processes over long periods of time this is less of an issue than the potential for stopping the data collection process because of incompatibilities of the data object and the database schema. The data collection process was automated using Linux System suite, effectively creating a service or daemon (a program running in the background) administered by the operating system. This allowed for minimal user

intervention, ensuring that stopped processes due to unexpected errors were all automatically handled and rebooted, allowing for almost completely unstopped data collection. This pipeline and automation process allowed for maximum reliability of data collection and data storage, with very little overhead in maintenance, which are key components in projects that require the collection of medium to large quantities of digital data. Scripts and daemon files are available in appendix A

Data ingestion pipeline

Due to the volume of the data used in this project, I was forced to consider in greater detail the stages of the data lifetime process, something that in projects with less data volume might be considered trivial. In the type of quantitative data analysis traditionally made in social science, the data file itself, usually a text file in the form of tabular data (e.g., csv or tsv) fits inside the working memory of consumer-level personal computers. This allows the processing programs (R, Python, STATA) to access it fast and reliably, and to perform modification and transformations over it in an efficient way. When data files are bigger than the available memory of the computer, traditional statistical packages struggle to work with the data, and only recently introduced integrations between big data frameworks, such as Apache Spark, and traditional analysis software such as RStudio or Python Pandas, allow to overcome this restriction. Still, such integration requires to set up two systems, namely the distributed database, and the front-end analytical tool. In this project I took a simpler approach. Raw data files were passed through a series of consecutive ingestion processes in order to reduce the size of the data and filter any unnecessary information contained in them.

Individual raw data files, as they were output from the document-based database system, were composed of approximately 20 to 30 millions of individual records (namely, tweets) in JSON format. The files themselves were of JSON newline delimited streams type. Essentially this is a text file in which each line is an individual Twitter JSON object (Twitter 2021b). The files were stored in a high-richness encoding system, namely UTF-8, to preserve special characters from languages other than English. This also preserves emojis present in the text of the tweets. For this project the choice of text encoding was less trivial than for most research projects. For the study of multilingual social media text, choosing a text encoding that can preserve the characters of each language is vital. On the Internet (websites and web browsers) UTF-8 multi-language encoding is the norm, but within analytical tools and common programming

languages used by the scientific community this is still not the case, and it is important to check at every stage of the process if multi-language encoding is being considered. Otherwise, there is the risk of losing important information when text is passed from one encoding to the other. Unfortunately, this comes with a downside. UTF-8 is a rather heavy encoding system compared with traditional ASCII, which in turn increases the size of the data files. Each data file had a raw size ranging from 120 to 180 gigabytes. This made it impossible to load the file in memory for processing, even when using the high memory High Performance Computing[2] facilities provided by the University of Leeds.

To solve the above mentioned problem, a line by line approach was used. In this strategy, since every newline of the raw data are a single self-contained JSON objects, it can be read, decoded, processed and then dumped into a file one by one. This strategy uses almost no memory, but relies heavily on the efficiency of the subroutine for reading and processing the raw data files, as well as the capacity of the hardware to read, process and write data to the storage devices. To solve the first issue, namely efficient subroutines, a compiled subroutine[3], written using the programming language Rust and the library Serde (Serde 2021), was used to read each file and process the data objects line by line. The use of Serde and Rust was necessary given the efficiency and performance in computation that Rust programming language offers compared to Python (Perkel 2020). This subroutine outputs a digested JSON file. This file was again of a "one line = one object" format, but each JSON object carried considerably fewer fields, greatly reducing the size of the digested file. The source code for the Rust subroutine, as well as an example of the digested JSON object can be found in the appendix A. In order to address the second issue, namely constrained performance due to hardware limitations, the whole ingestion and downsizing step was carried out within the University of Leeds High Performance Computing (HPC) facilities administered by the Research Computing Group of the University of Leeds, which allowed me to have access to high speed storage devices and high speed computer nodes, which greatly increased the speed of the ingestion and processing. Performing this process on regular consumer-level hardware would have increased the computation time to the orders of

[2] A High Performance Computer is a purposely built computation facility, often involving many nodes (a node is similar to what is usually consider a personal computer) and high-performance supporting hardware (e.g., very fast storage drives, very fast CPUs, large amounts of working memory, etc.)

[3] Programming languages can be divided into compiled and interpreted languages. As a general rule, subroutines written in compiled languages are more resource efficient and faster than the same subroutine written in an interpreted language. This performance differential is due to several factors, but a very important one is that in compiled languages, many decisions had to be taken upfront by the programmer (e.g. the type of the data and type of variables), while in interpreted languages those decision are automated in favour of usability but in detriment of the amount of operations to be handled by the interpreter, and hence in detriment of performance.

several days per file, while on the HPC cluster it took several hours per file.

4.2 Methodological Approach Selection

In this project, I will use the text of the tweets to infer its emotional content. This task falls in the field of Natural Language Processing and text mining. Although, tweet objects contain URLs to all the multimedia resources associated with each tweet (e.g., videos, images) that one potentially could use to conduct emotional inference, I will only use the text of the tweets, as inferring emotions from images and video is considerably more demanding, both methodologically and computationally.

Natural Language Processing involves the automatic processing and discovery of the underlying patterns in natural language data, this generally being the spoken and written language of human beings. This can be done using data from vocal sounds or from text. I will use text, and while text is certainly more mediated than speech, it does still pose several challenges. Text data are high-dimensional. Length, casing, style, language, character set, etc. are all features of text data, and online text is usually very unstructured and noisy (Yang and Pedersen 1997). In order to gain knowledge from these noisy and high-dimensional data via algorithmic methods first we need to narrow down the nature of the problem that we want to tackle. Since the success of this project relies on correctly identifying the emotional content of a string of text with the highest accuracy possible —as this is the main independent variable— a significant portion of my analytical efforts must be focused in determining the best technique to identify said emotional content. This task falls into the subdomain of text classification. Over the years, a variety of techniques have been developed to perform text classification. Roughly speaking, these methods can be organised in two big categories: Lexicon-based methods; and supervised machine learning methods.

In order to determine the most suitable technique for text classification to be used in this project, I have selected a pool of methods which are currently popular in text classification tasks. In the following section I will review the general characteristics of each family of methods and I will provide an in-detail explanation of the specific techniques. In Section 5.1 I will provide a detailed comparison of the performance of each classification technique on which basis I will choose the best one for the final classification of the data.

4.2.1 Training and testing data

Several of the algorithms to be tested need "labelled data." In this case this translates to sentences to which a label, corresponding to their most prominent emotion, has been assigned. All of the classification methods also require "testing data," which is usually extracted from the same data source of the original labelled data, by random sampling a significantly smaller subset of it, leaving the majority of the data as "training data." There is no rule-of-thumb or general standard to define how much labelled data is enough to create "good" training and testing datasets. The general practice is to amass as much labelled data as is possible for the task at hand. The sources of labelled data are basically two: purposely built datasets, and general purpose third-party datasets. Purposely built datasets are a good option when the task is very specific or the data itself are difficult to obtain. However, this comes with the trade-off of being much more costly for the researchers, because of the time and resources it takes to curate and annotate each individual data point. Third-party, general purpose datasets, on the other hand, are less specific but many are released under free user licences, which greatly reduces the costs for the researcher. Sadly, this comes with the risk of reduction in the quality of the data and/or the annotations. The flaws in quality can be of several kinds. Bad annotation is certainly the most glaring for a classification task like the one in this dissertation. However, that flaw is usually evident once the models trained on such third-party training data are tested on actual data of interest. I will use general purpose, third-party datasets, as the advantages in terms of cost reduction out-weights the risk of bad quality data annotation.

4.2.2 Training data labelling

The theoretical reasons to select anger, fear, sadness and joy, as the emotions to be used in the analysis are outlined in Section 2.3.1 of Chapter 2. Additionally, these four emotions resulted to be the most common labels across the training datasets in both languages, English and Spanish. The reason for this is that many of the training datasets based their labelling process in the work of Ekman and Friesen (1971), which includes anger, fear, sadness and joy as a subset of his basic emotions taxonomy.

English dataset

The English training and testing dataset used in this research was constructed from a variety of third-party "emotion-detection" datasets in order to increase the size and variability of the

final composite dataset:

1. "Emotion Classification NLP" (Tripathi 2022): A dataset composed of around 7000 tweets, annotated with four emotions (anger, fear, sadness, joy). It was released under Creative Commons 0 licence for public use, free of any conditions, by Anjaneya Tripathi.

2. "Emotions dataset for NLP" by Saravia et al. (2018): A dataset composed of 20,000 sentences annotated with six emotions (anger, fear, sadness, joy, love and surprise). Released under Creative Commons CC BY-SA 4.0.

3. "SMILE Twitter Emotion Dataset" by Wang et al. (2016): A dataset containing 3,085 tweets, annotated with five emotions (anger, disgust, happiness, surprise and sadness). Released under Creative Commons CC BY 3.0.

In cases where the datasets contained data points labelled with emotions not to be used by this project (e.g., disgust, surprise, love), those instances were discarded from the final ensemble of training and testing data.

In order to perform an adequate training process for each of the tested algorithms, this composite datasets must be divided, using random sampling, into a training set (80% of the original dataset), which will be used to train iteratively each of the classification algorithms and a testing set (20% of the original dataset), which will be used as a gold standard to calculate the accuracy of the already trained algorithms (for more details see Section 4.2). Ideally, the training and testing datasets must retain as much as possible the characteristics of the original dataset. In my case, I tested this using character length distribution (see Figures 4.1 and 4.2).

The final English composite dataset contains 22,818 labelled data points, and its final labels distribution is the following:

Labels	Counts	Train Counts	Test Counts
joy	8799	7075	1724
sadness	6687	5315	1372
anger	3705	2966	739
fear	3627	2898	729

Table 4.4: Distribution of data points per label in English composite dataset.

As can be seen in Table 4.4, the label distribution is unbalanced, with a sizeable majority of data points being classified with "joy" and "sadness." This lack of balance however is still "mild"

(Google 2022). In the current machine learning literature, unbalanced datasets are considered a serious problem when the proportion between the majority and the minority classes is close to 100:1 (He and Garcia 2009). This is not the case here, hence no balancing correction across classes was performed.

The distribution of character length for the original composite and the training and testing English datasets can be seen in Figure 4.1

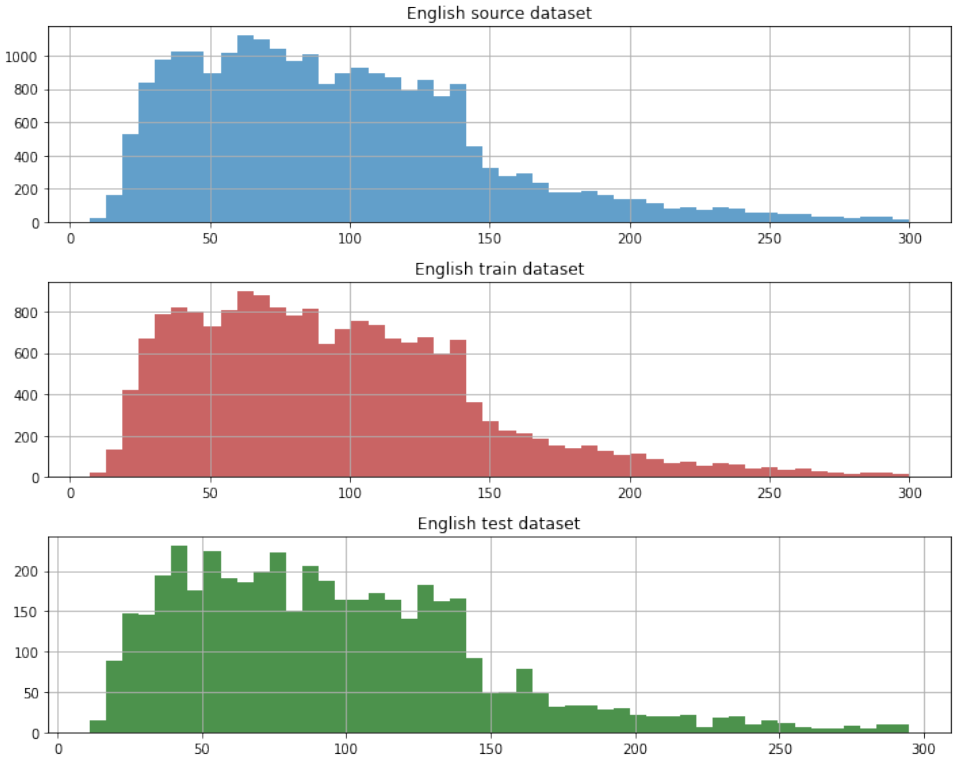

Figure 4.1: Distribution of character length for English source dataset (top), training dataset (middle), and test dataset (bottom).

As can be seen in Figure 4.1 and in Table 4.4, the English testing and training datasets do not differ significantly from their source data. We can also notice that the character length distribution in the data, although not all being tweets, does fall into the ballpark of the 280 length cap on Twitter, which shows the suitability of the data for building a model for Twitter data.

Spanish dataset

Natural language processing in non-English contexts have additional constraints. Given the evident dominance of the English language in scientific literature and technological development,

English has become the de facto lingua franca in science. While this comes with many advantages (e.g., improved international collaboration, simpler dissemination, etc.), it does have some non-trivial consequences. In particular, for the field of NLP, it means that early developments and resources are concentrated on English, in detriment to any other languages. These developments and resources include, of course, the amount and variety of annotated text data for any natural language processing task (Djatmiko et al. 2019; Kaity and Balakrishnan 2020). This constraint made the search for large and varied emotion detection Spanish dataset significantly more difficult than for English. Gladly, SemEval —the prestigious ongoing series of computational semantic evaluations— in its 2018 version, put together a series of emotion detection competitions along with several multilingual emotion-annotated tweets. This task was specifically designed to detect emotions in English, Arabic and Spanish tweets using the same four emotions as in this project (anger, fear, sadness and joy).

The datasets used in the Spanish version all come from the compilation made by Mohammad, Bravo-Marquez, et al. (2018) for the 2018 version of the SemEval Competition (https://semeval.github.io/), Task 1.

The Spanish source dataset used here is a copy of the one provided by Mohammad, Bravo-Marquez, et al. (2018) and contains 8541 labelled data points. The Spanish source, training (80% of the original dataset) and testing (20% of the original dataset) dataset all have a very similar label distribution as can be seen in Table 4.5.

Labels	Source Counts	Train Counts	Test Count
joy	2975	2404	571
sadness	2181	1721	460
anger	2080	1654	426
fear	1305	1053	252

Table 4.5: Distribution of data points per label in Spanish source, training and test datasets.

Again, there is a "mild" imbalance in the dataset's classes, but this does not reach problematic levels. Additionally, as can be verified in Figure 4.2 all Spanish datasets have a very similar character length distribution, with a hard stop at 280 characters, given this is a collection of tweets only.

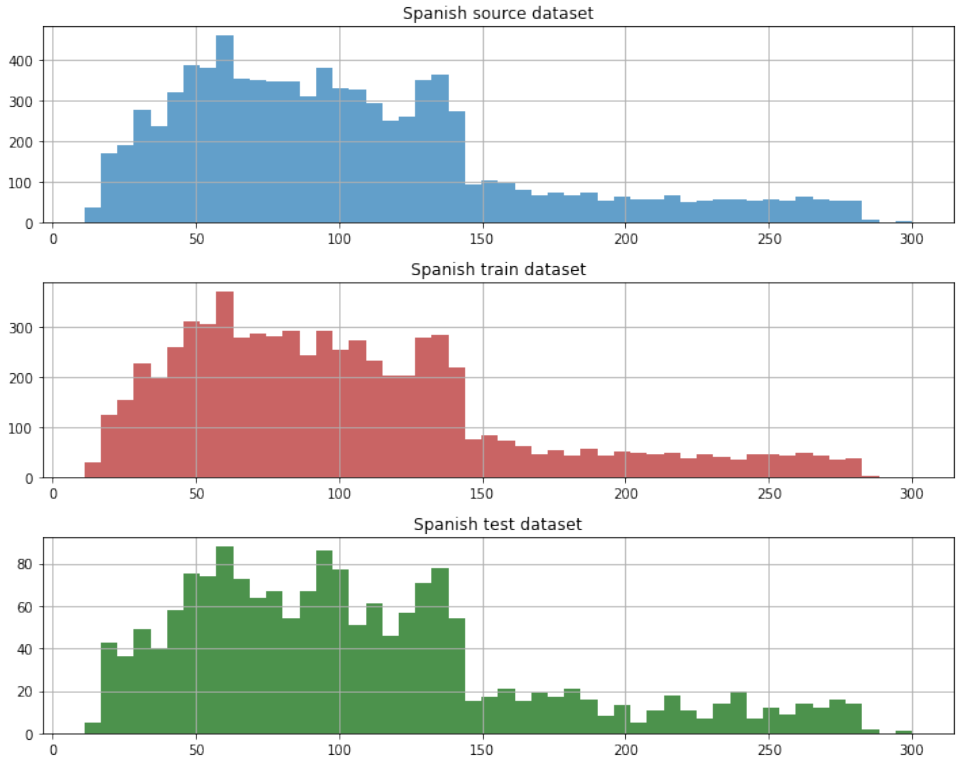

Figure 4.2: Distribution of character length for Spanish source dataset (top), training dataset (middle), and test dataset (bottom).

Training data cleaning

Only essential data cleaning was performed in order to preserve the maximum amount of information in the training data. Notably, only URLs and user tags were removed from all datasets before the training process of the classifiers, in order to avoid spurious relations with specific users or websites. They were replaced by a "[URL]" or "[USER]" placeholder in order to keep as much as possible the syntactic structure of the text. No other sequence or special characters were removed, and no lower-casing was enforced.

4.2.3 Evaluation metrics

When evaluating the performance of machine learning methods it is necessary first to consider the nature of the task at hand. Usually this performance evaluation considers the accuracy of the classification and the overall efficiency of the algorithms, as it is necessary to find a balance between performance and consumed resources. However, for this project, being a one-off multi-class classification task, we are interested in finding the method (or methods) that can classify with the greatest accuracy, regardless of the consumed resources it takes to perform the

classification. Following this logic, The metrics to be used to compare the different algorithms with respect to their accuracy are:

- **Precision score**: Defined as the ratio of true positive classifications over the sum of true positive and false positive classifications. High values of this score imply the classifier correctly identifies the true positive cases, but it does not say anything about the false negatives. A very precise classifier might be only picking up the most evident cases.
- **Recall score**: Defined as the ratio of true positive classifications over the sum of true positive and false negative classifications. This score provides information about the capacity of the classifier to pick-up all the relevant cases, regardless if it makes a lot of mistakes.
- **F1 score**: The F1 score is a measure of accuracy of the classification, which combines the precision and recall scores. It corresponds to the harmonic mean between precision and recall, and when those are similar it approximates to the average between the two. It is calculated as :

$$F_1 = \frac{2}{recall^{-1} + precision^{-1}} \tag{4.1}$$

  The F1 score provides an overall evaluation of the performance of the classifiers. The higher the F1 score is, the better the classifier is. In addition to the raw F1 score, macro and weighed versions of the F1 score will also be provided. These two additional versions of the F1 score provide an average score across all of the classes the classifier works on, and are recommended to examine when the number of classes is above two (in this case is four). Macro F1 score provides an average across the F1 score of all classes, while Weighted F1 score provides a weighted average across the F1 score of all classes.

4.2.4 Tested Methods

Lexicon-based methods

Lexicon-based methods for sentiment analysis of social media text have been used in social and political research for a while now (Pak and Paroubek 2010; Verma et al. 2011; Kouloumpis et al. 2011; Nielsen 2011; Agarwal et al. 2011; Sen et al. 2015; Bakshi et al. 2016). In general terms this technique involves giving the document a score on a polarity scale ranging from negative to positive; or a completely categorical classification, like, for example, an emotion. The way this is

achieved is by comparing the words in the documents with a lexicon, in which certain sentiment words have a score on this polarity scale, or labels classifying them as negative, positive or neutral. Later these scores are aggregated at the levels of analysis chosen by the researcher, usually sentence or document level, finally obtaining a general sentiment score.

Lexicons can be thought of as tables with a column containing a long list of usually thousands unambiguous sentiment-loaded words (or unigrams), a column with the corresponding numeric sentiment score or sentiment category, and sometimes an optional column containing the intensity of the sentiment in the form of a numeric score (for the cases in which the sentiment is represented as a collection of features rather than a single continuous score) (Mohammad 2017; Khoo and Johnkhan 2018). The process by which the lexicons are created and annotated can involve human annotation or automatic annotation (Mohammad 2017). This process can also be extrapolated to create Emotion Lexicons, in which a word can be labelled as expressing one or multiple specific emotions. This process assumes that words can convey more than one emotion, but each of them must be given a certain weight. This usually takes the form of a table in which words have different scores for each of the emotions in the table.

The main advantage of the lexicon-based approach is its low computational cost. Beyond the tokenization and, possibly, cleaning process of the documents, there are little calculations done by the computer, which in theory makes it more suitable for large quantities of text to be analysed. The disadvantages of using lexicon methods are varied. The first one is related to its construction. Significant amount of resources are invested in creating an annotated lexicon. Usually human annotation is preferred, since automatic methods might be less accurate, but also increases the time and monetary resources to be invested in the process (Mohammad 2017; Gatti et al. 2016). This forces lexicons to be created with a strategic goal in mind, usually targeting highly used languages and tuned to general purpose use, leaving languages other than English or Spanish usually without support (Wali et al. 2020). Another big disadvantage is its inability to weigh the context of the sentiment-loaded words. Since these methods involve comparing the words of the document with the words in the lexicon, all words not present in the latter are essentially ignored. This means that modifier words like *very* or *extremely* are not considered when calculating the sentiment of a sentence/document. This leads to the situation in which the sentences *I'm sad* and *I'm very sad* will have the same sentiment score (Mohammad 2017). Although there have been attempts to address this issue by adding an additional table

in the lexicon which can gauge the effect of modifier words (Thelwall et al. 2010), this solution has not been generalised to other lexicon-based libraries. The order in which the words appear in the document/sentence is also not considered by lexicon approaches as well as negations. For example, *I don't feel good* and *I feel good* would be both evaluated as "positive." This has led to the development of lexicons labelling not just one but a group of words (bigrams or trigrams) (Mohammad 2017). Although this might address to some extent the drawbacks of the method, it also substantially increases the amount of entities to be labelled, making this solution unsustainable in the long run.

Despite these disadvantages, lexicon-based approaches are still in use today and are a very popular natural language processing technique, especially if there are no resources to implement more sophisticated methods. In this project one lexicon method was evaluated, the NRC Emotion Lexicon.

The NRC Emotion Lexicon

The NRC Emotion Lexicon, also *EmoLex* (Mohammad and Turney 2010), is an English emotion lexicon containing 14,182 unigrams (words) and around 25,000 senses, senses being different contextual variations of the words, labelled for 8 emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and 2 sentiments (positive and negative). The annotation process was performed by a Human Intelligence Task (HIT) on Amazon MTurk, and it is provided free of cost for academic research. Being a general purpose lexicon its application to short social media text falls within its scope of application. Its performance with Twitter data is reviewed in Section 5.1.1.

Lexicon Methods Testing Setup

Lexicon-based methods do not have hyper-parameters to be tweaked. This does not mean that there is nothing to adjust previous to the classification process. Lexicon-based methods are built with general purpose in mind, and it is common for them to include a large variety of emotions in their dictionaries. Many of those emotions can be related or are very close both in meaning and in the numerical values it can take in the lexicon's dictionary. This poses the question of what is considered a correct classification. To solve this issue there is no other way than to use arbitrary criteria. So, a correct classification of our lexicon-based methods will be a correct identification of the emotion **if** that emotion is present in the dictionary. If the true emotion

label is not present in the dictionary, that class (or classes) will be excluded from the evaluation. An argument could be made to use a "close enough" emotion as a correct classification (e.g., "hate" might be considered close enough to "anger"). However, that might pose the problem of determining when an emotion is "close enough," which is a discussion outside the scope of this study. However, the same emotion spelled differently or in a different conjugation (e.g. "anger" vs "angry") are going to be considered the same.

Supervised Machine Learning for text classification

Machine learning is a field focused on finding ways by which computers can gain knowledge about reality by providing them with data to learn from (Jordan and Mitchell 2015). This learning process implies the use of an algorithm that can inspect the data and infer some generalizable patterns from it. These algorithms can be constructed upon assumptions of their parameters based on probability distributions (like linear or logistic regression) or be assumption-free about their parameters like k-Nearest-Neighbours. Additionally, machine learning methods can be divided into supervised and unsupervised algorithms.

Supervised classification algorithms (or supervised learning classification algorithms) take that name because the data passed through the algorithms contains a set of features as well as the label for each sample the algorithm needs to learn to distinguish between sample classes (Jordan and Mitchell 2015). Let us imagine that we have the task of developing a supervised machine learning classifier to distinguish samples of a flower between their three most well known subspecies. To do so we will collect data from the three subspecies of the flower and for each we will have four features: sepal length, sepal width, petal length, petal width. Additionally, for all the samples, we have their corresponding sub-species. Then, the classifier will look at each instance in the data and will adjust its internal parameters to weight the features provided in order to make the most optimal possible distinction among the labels (in this case, the three subspecies). it is like learning by trial and error. This process requires to separate the data into a training set, which usually contains the majority of the data, and a testing set, in which the classifier will attempt to correctly predict the labels on unseen data. Based on the accuracy of the predictions with the testing set, a researcher can evaluate the performance of the trained classifier.

Supervised learning algorithms require labelled data to be trained and tested on. For relatively

simple tasks, the data collection process already includes the labels for each of the samples, like in our example with the flowers above, but for complex task, like recognising a latent characteristics (like the emotions present in a sentence, for example) a separate annotation process of the collected data is necessary. Usually this annotation is done by trained or untrained human annotators or by experts in the field. In this case the machine learning algorithm is learning to make the classification that an annotator would have made, thus saving time, rather than discovering objective truth. Depending on the complexity of the task, annotation can be very costly, as larger quantities of labelled data are needed to train a reliable classifier. Despite these drawbacks, supervised machine learning algorithms are very popular given their accuracy and versatility (Jordan and Mitchell 2015). I will now explore a range of supervised classification algorithms, which were considered here as an approach for identifying emotions in tweets.

Support Vector Machines

Support vector machines (SVMs) are a popular supervised machine learning algorithm that can be applied to tasks of classification and regression. Their modern form was first proposed by Boser et al. (1992) and since then it has been used extensively in text classification tasks (Tong and Koller 2002) with relatively good results in comparison with other popular machine learning methods for classification like Naive Bayes or k-nearest neighbours (Kadhim 2019). This makes it a good choice for the emotion classification task in this project.

The premise by which SVMs work is simple in principle. Given a dataset in which we have at least one feature and a class label with two or more classes, SVMs will create a parameter space using the provided features. It will place the observations in the space according to their respective values in said features, and will find the boundary of separation between the classes.

Coming back to our example with the flowers, Figure 4.3 shows a simple example in which the SVM finds the optimal boundary between two different sub-species of Iris flowers using the features of sepal length and sepal width. As can be seen in Figure 4.3, there is one misclassified red dot appearing in the blue group cluster. One could ask why does the classifier not simply draw the boundary closer to the blue dots, allowing for perfect classification. This apparent mistake of the classifier will help us to better understand its inner workings.

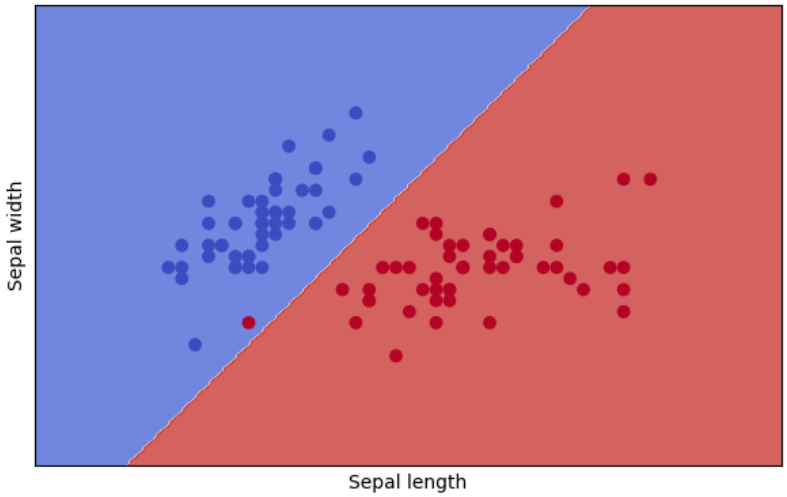

Figure 4.3: Support Vector Machine example. The line separating the points (two sub-species of Iris flowers) represents the optimal boundary between the classes. Source: Personal collection.

Support vector machines work with something called *Soft Margins* classifiers or *Support Vector* classifiers. These classifiers allow for misclassifications. A classifier that finds the best boundary to separate the classes and allows for no misclassification is called a *Maximum Margin* classifier. However, maximum margin classifiers are very sensitive to outliers, like the case of our red dot. If we were to move our boundary in Figure 4.3 closer to the blue dots, as a maximum margin classifier would do, we might classify correctly all of our training data, but we might miss-classify new unseen data in the future that are actually part of the blue group, but a little out of our very strict margin. In short, maximum margin classifiers generalise poorly, and thus, soft margin classifiers, or support vector classifiers are preferred. In Figure 4.4 we can see the same data as in Figure 4.3 but now showing the soft margins dashed lines inside which misclassification is allowed. We also note that there are data points over which the soft margins lines pass right on top. These data points are known as the *support vectors*.

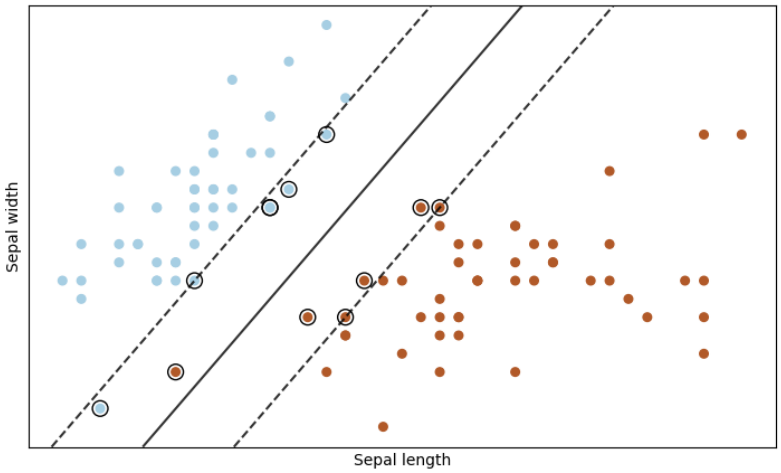

Figure 4.4: Support Vector Machine example. The dashed lines represent the soft margins inside which misclassification is tolerated. Source: Personal collection.

The intuitions described above can be mathematically formalised using the following notation.

Let's assume that we have the following training dataset of n data points:

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n) \tag{4.2}$$

Where y_i can take values only of 1 or -1, representing the class to which each $\mathbf{x}_i$ belongs to. Each $\mathbf{x}_i$ is a p-dimensional vector of real numbers. This vector essentially represents the values each data point can take in the p numbers of features that were measured. The goal of the SMV is to find the hyperplane (a hyperplane in $> p$ dimensions, which then becomes a complex surface projected down to p dimensions) that perfectly separates the two groups. A hyperplane perfectly separating our two groups can be written as a set of data points $\mathbf{x}$ satisfying:

$$\mathbf{w}^T\mathbf{x} - b = 0, \tag{4.3}$$

Where $\mathbf{w}$ is the normal vector or magnitude of the hyperplane (the perpendicular line heading away from our hyperplane, a concept useful to calculate Euclidean distances), and b is an offset of the hyperplane from the origin along the normal vector $\mathbf{w}$. Similarly, we define the hard margins of the separation using the following hyperplanes:

$$\mathbf{w}^T\mathbf{x} - b = 1, \text{ if } y_i = 1 \tag{4.4}$$

Where 4.4 is the hyperplane above which every point is of class 1,

$$\mathbf{w}^T\mathbf{x} - b = -1, \text{ if } y_i = -1 \tag{4.5}$$

And where 4.5 is the hyperplane below which every point is of class -1.

The distance between hyperplane 4.4 and hyperplane 4.5 is given by $\frac{2}{\|\mathbf{w}\|}$. So in order to find the maximum distance between this two, we need to minimise the value of $\|\mathbf{w}\|$. Figure 4.5 shows now the hard margins in dashed lines.

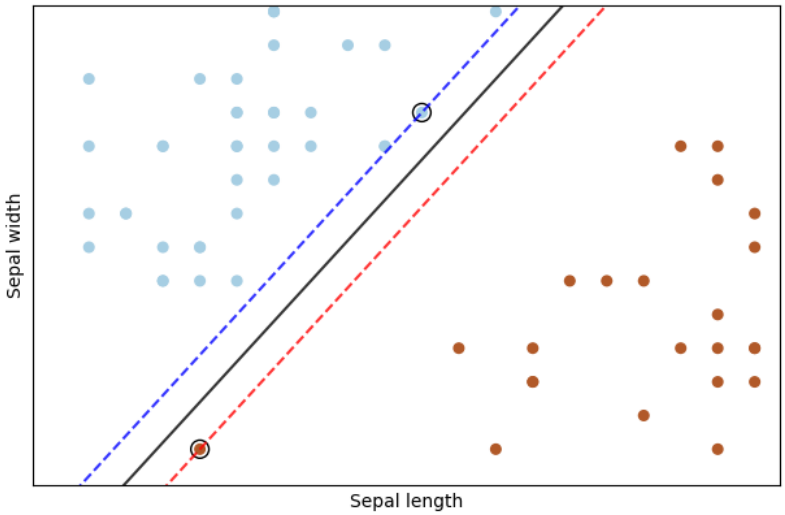

Figure 4.5: Support Vector Machine hard margins. The hyperplane described in equation 4.3 is represented by the black solid line, the hyperplane described in equation 4.4 is represented by the dashed blue line, the hyperplane described in equation 4.5 is represented by the dashed red line, and the distance between the dashed red and blue lines is described by $\frac{2}{\|\mathbf{w}\|}$. Source: Personal collection.

Equations 4.4 and 4.5 can also be rewritten in one expression as follows:

$$y_i(\mathbf{w}^T\mathbf{x} - b) \geq 1, \text{ for } i = 1, \ldots, n. \tag{4.6}$$

Then, the whole minimisation problem can be expressed as:

$$\text{minimise } \|\mathbf{w}\| \text{ conditional to } y_i(\mathbf{w}^T\mathbf{x} - b) \geq 1, \text{ for } i = 1, \ldots, n. \tag{4.7}$$

If we stop at equation 4.7 we obtain a hard margins classifier. But, as was mentioned above, hard margins classifiers generalise poorly to new data. To solve this, we can modify this problem by adding a function that allows for flexibility. This function is called the *hinge loss function*, which penalises over-fitting, encouraging the boundaries not to lie too close to one class (for more details see Smola and Schölkopf (2004)). Applied to our minimisation problem, the result is:

$$\max(0, 1 - y_i(\mathbf{w}^T\mathbf{x} - b)) \tag{4.8}$$

If the condition in equation 4.7 is fulfilled, and every point is correctly classified, the hinge loss function outputs 0. If not, the function produces values proportional to the distance of each point to the margin. This function can then be used as a weight for our minimisation problem as follows:

$$\min \lambda\|\mathbf{w}\|^2 + \left[\frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(\mathbf{w}^T\mathbf{x} - b))\right] \tag{4.9}$$

Where $\lambda > 0$ is a hyper-parameter representing the trade-off of increasing the margin size and misclassifying data points.

In simple terms, equation 4.9 will attempt to find values of $\mathbf{w}$ and b that yield the lowest possible value of $\|\mathbf{w}\|$ given the data vector $\mathbf{x}$ for each outcome point y_i. Then, if any misclassification occurs, it will use the hinge loss function to give a proportional value to it depending of how far the misclassified point is to its real class, to then average all of the misclassifications and add them as a penalty to the final value of $\|\mathbf{w}\|$, that we need to minimise. This is the *primal* problem of the classification process in SVMs (for more details see Smola and Schölkopf (2004)), meaning the primary problem that we want to solve in SVMs in order to obtain the optimal solution to our classification.

Finally, while support vector classifiers are useful in many situations, there are cases in which the classes are not separable in their raw feature space (e.g., the feature space formed by the petal

width and petal length in Figures 4.3 and 4.4). To solve this problem, we can create additional dimensions for our feature space, for example, by squaring the raw features. This transformation is performed by what is known as the *kernel* function, and the combined operation of finding the optimal support vectors in the optimal feature space derived from the data using the kernel function, is what is known as *Support Vector Machine*.

SVMs can be used in many fields as long as the features can be represented in some numerical form. When the goal is to classify text into some categories, the text itself undergoes a transformation process, so its properties (e.g., words, length, position of the words, etc.) can be represented in the form of numerical vectors. Then these vectors are used as the features to train the SVM classifier.

Support Vector Machines Testing Setup

Support vector machines have a number of "flavours," which are determined by the base kernel to be used (for more details see Smola and Schölkopf (2004)). In my case I will test two SVMs kernels linear and Gaussian radial function. These two kernels cover most of the use cases of SVMs. A linear kernel assumes that the classification is a problem that can be solved by tracing a line (or several lines) in the feature space in order to separate the classes, or, in other words, that the classes are "linearly separable." The Gaussian radial function kernel, on the other hand, assumes that the problem is not linearly separable and attempts to find a non-linear separation between the classes, by, for example, using areas or volumes to separate the classes. In my case, I do not have prior information suggesting that the problem of emotion detection and classification in text can be either linearly or nonlinearly separable, so it makes sense to try both options. All other hyper-parameters were left at their default values.

Naive Bayes Classifiers

Naive Bayes Classifiers are an application of the Bayes theorem to a classification problem. In essence, by applying Bayes theorem to the problem, the classifier is a method to calculate the conditional probability of the occurrence of the class given the information seen at the training stage, and the information provided by the set of features of the new data to be classified (Zhang 2004). To understand this better, I will briefly review the Bayes theorem.

Bayes law is a probability theorem. The probability of an event A is commonly denoted as

$P(A)$ and is usually defined, in Bayesian terms, as the probability that A will occur as a state of belief. It can be estimated by taking the number of desired outcomes, divided by the total number of all outcomes (e.g., number of heads of the total number of coin tosses). Conditional probability of an event A given B, on the other hand, is denoted $P(A|B)$ and is defined as the probability of A occurring given that B already occurred, and can be calculated as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{4.10}$$

For simplicity, the above equation can also be rearranged as follows:

$$P(A \cap B) = P(A \mid B)P(B) \tag{4.11}$$

Where $P(A \cap B)$ is the probability that both events A and B occur. If A and B are deemed to be independent events, it can be calculated as $P(A)P(B)$, meaning, the multiplication of the independent probabilities of the events A and B. This leads to the conclusion that $P(A \cap B) = P(B \cap A)$ since $P(A)P(B) = P(B)P(A)$, given the commutative property of multiplication. However, we must note that $P(A \mid B) \neq P(B \mid A)$. Equation 4.10 describing $P(A \mid B)$ is not the same as the equation 4.12 describing $P(B \mid A)$, and it is a common mistake to assume that those are equivalent.

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \tag{4.12}$$

In equation 4.11 we can see that $P(A \cap B) = P(A \mid B)P(B)$ and using the same equation we can also assume that $P(B \cap A) = P(B \mid A)P(A)$. Now, given that $P(A \cap B) = P(B \cap A)$, and given what we learned from equation 4.11, we can now say that:

$$P(A \mid B)P(B) = P(B \mid A)P(A) \tag{4.13}$$

And if we rearrange equation 4.13 to leave only one term on the left side we arrive at the Bayes law (Bayes 1763), which describes the relationship between $P(A \mid B)$ and $P(B \mid A)$:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{4.14}$$

The mere description of Bayes Law in the previous section does not necessarily help us to understand how or why such a theorem can or should be used in a classification algorithm. To do so, we must first take a step back and reflect what it is that a classifier does. Classification algorithms attempt to make an informed guess about the class to which a new data point belongs. This informed guess is based, in the case of supervised algorithms, on training data. However, the methods through which each algorithm makes the distinction vary widely. Support vector machines (see 4.2.4), for example, make this distinction by quite literally drawing lines in the feature space that best separate the classes. Other algorithms might use simple binary rules, like decision trees (Murthy 1998). Given this, we can also think of other ways in which an algorithm can make informed guesses. One of them is by using probabilities and probability distributions. The algorithm will attempt to calculate the probability of a new data point belonging to a class, given the proportions of each class in the training data, and giving the feature's distribution in the training data. This can be expressed in mathematical notation as:

$$p(C_k \mid \boldsymbol{x}) = \frac{p(C_k)p(\boldsymbol{x} \mid C_k)}{p(\boldsymbol{x})} \tag{4.15}$$

In equation 4.15 we use the Bayes Law to calculate the probability p of occurrence of each class C_k among all K classes, given that we have observed the vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ of n features, which represent our training data. When all the probabilities for each class are calculated, we can simply take the most probable class as our classification prediction. This procedure is known as *maximum a posteriori*. The different versions of a Naive Bayes result mainly from the type of distribution the features in the data vector $\boldsymbol{x}$ have. If the features are continuous normally-distributed variables, a Gaussian Naive Bayes is more appropriate. On the other hand, if the features are discrete, a Multinomial Naive Bayes might be a better option.

Naive Bayes Classifiers Testing Setup

Information on the features' type of distribution is crucial to determine which type of Naive Bayes we must use. This means that it is technically possible to have flavours of a Naive Bayes classifier for as many types of distributions as the features can take (e.g., Poisson Naive

Bayes, Gamma Naive Bayes, Bernoulli Naive Bayes, etc.). Here we are partially limited to the options provided by `scikit-learn`: "Gaussian," "Multinomial," "Complement," "Bernoulli" and "Categorical." Unlike with the case of the class separation in SVMs, here we do have previous relevant assumptions and information to help us to select the most suitable algorithm. Our feature space is composed of discrete units: word counts. This information enables us to discard the Gaussian Naive Bayes classifiers since this version is meant to be used with continuous parameters spaces. Additionally, a test conducted by McCallum, Nigam, et al. (1998) showed that the most appropriate Naive Bayes models for text classification are the multinomial and Bernoulli based classifiers. The Complement Naive Bayes classifier is a variant of the multinomial classifiers adjusted for class imbalance. Given that both our training datasets present some mild imbalance in their class counts, I will include it as well in the test group. Finally, all hyper-parameters of the classifiers were left to their default values.

Deep learning methods for text classification

Deep learning has taken the field of machine learning by storm in the last decade. Deep learning models have provided applicable solutions to problems that previous methods struggled with for decades (LeCun et al. 2015). Thanks to the immense amounts of data produced today on the Internet, entities (individuals, companies or research centres) have found ways to take advantage of this sea of information to train deep learning algorithms that can classify medical images, drive cars, or recognise speech (LeCun et al. 2015). And the field of Natural Language Processing has benefited greatly from these methodologies. Since the introduction of the *transformers* architecture by Vaswani et al. (2017) many problems in which previous architectures, like Recurrent Neural Networks (see 4.2.4) or Long Short-Term Memory RNNs (see 4.2.4) had been struggling, now have easy to use solutions, using only a fraction of the data required to train previous deep learning models. But in order to understand these advances better it is necessary first to understand the basic algorithm on which all forms of deep learning are based: The Multilayer Perceptron.

The Multilayer Perceptron

The multilayer perceptron (MLP) is the building block of all deep learning methods and one of the most simple forms of artificial neural networks. The inspiration for MLPs comes loosely

from what was thought to be the functioning of neurons inside the brain (Rosenblatt 1958)[4]. In that metaphor, a set of neurons is able to capture some information/data, that then is passed to a set of hidden, internal neurons who learn some of the characteristics of this information and are able to form generalizable abstractions about it. Finally, this abstraction saved in the neurons, can be used to recognise or classify unseen information/data.

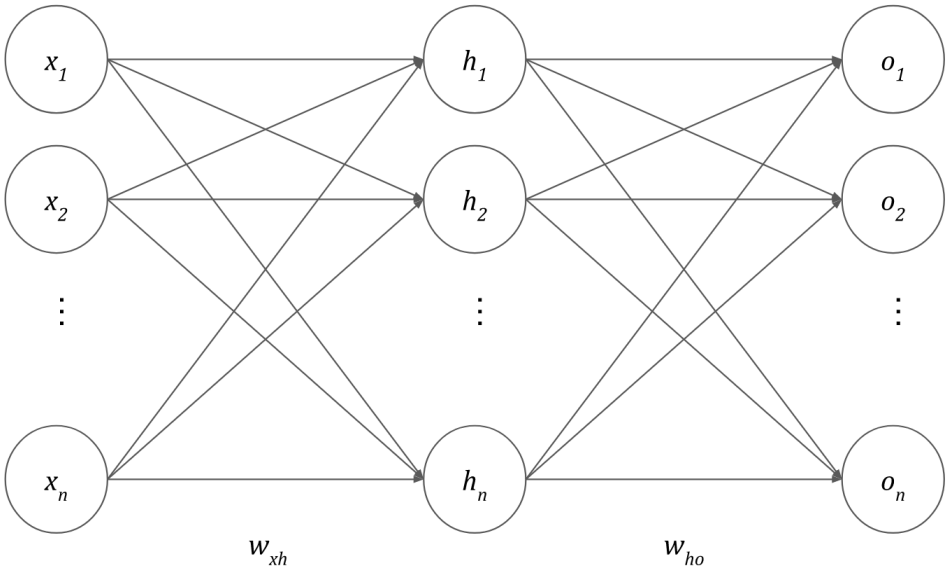

Figure 4.6: Multilayer Perceptron diagram showing the input layers x, hidden layer h, output layer o and the weights w_{xh} and w_{ho}. Source: Personal collection.

By a process known as *feed-forward* (Hornik et al. 1989) MLPs can receive input data and perform classification tasks. Using Figure 4.6 we can describe step by step the process of feed-forward (named that way because the values are passed through the network in a unidirectional form). x represents the input nodes coming from numerical features in our dataset (e.g., sepal length and width if we follow our previous example). The h nodes at the centre represent the hidden layer of the MLPs. Each of these nodes receives a weighted sum of all the input nodes:

[4]Neurons in the brain do have an threshold for activation that is similar, in a superficial level, to the activation function in an artificial neural network (ANN). However, real neurons exhibit plasticity: the ability to change the connections by strengthening them, severing them or diminishing them, depending on a variety of factors, which is a characteristic artificial neural networks notably lack. While ANNs can change the strength of their weights, its structure is fixed in the design step, and never changes after that. In addition to that, being a cell, real neurons have several other ways to transfer information between them, involving direct passive chemical communication (e.g., releasing hormones), or active transport of molecules via vesicles. Finally, there is the problem of substrate. Real neurons **are** the structure that makes the *computations*. In the case of ANNs, the substrate is the GPUs or CPUs, while the neurons are just an abstraction in software. This difference could be thought as too obvious to be worthy of any discussion (ANNs are just modelling tools after all), however it does carry strong consequences. Parallelisation capacity and energy consumption, for example, are much more efficient in real neurons in great part because they are the structure doing the work.

$$\begin{aligned} &\sum_{i=1}^{n} (x_i \times w_{xhi}) \text{, for each } (x_i,\, h_i) \text{ pairs} \\ &\sum_{i=1}^{n} (h_i \times w_{hoi}) \text{, for each } (h_i,\, o_i) \text{ pairs} \end{aligned} \tag{4.16}$$

After receiving the weighted sum from all their input nodes, the hidden nodes *compute* whether to become activated or not. This decision is performed by an activation function, usually of a sigmoid form. The goal of the activation is to provide a sudden change in the output value once a certain threshold in the input value is crossed, this is why functions with sigmoid, S like curves are preferred (e.g., the logistic function or the hyperbolic tangent function). In its most simple form, the complete feed-forward algorithm with a logistic sigmoid function: sigmoid $= \sigma = \frac{1}{1+e^{-x}}$, and using matrix multiplication to reduce 4.16 to a single operation, can be represented as follows:

$$\begin{aligned} \mathbf{h} &= \sigma(\mathbf{w}_{xh}\mathbf{x}) \\ \mathbf{o} &= \sigma(\mathbf{w}_{ho}\mathbf{h}) \end{aligned} \tag{4.17}$$

Where $\mathbf{w}_{xh}$ represents the matrix of weights between the layer x and layer h; $\mathbf{w}_{ho}$ represents the matrix of weights between the layer h and layer o; and $\mathbf{x}$, $\mathbf{h}$ and $\mathbf{o}$ represents the vector of values of each nodes of the layers x, h and o respectively.

It is important to note that the number of nodes in the output layer o must be the same as the number of the classes we wish to classify. These classes need to be available in our dataset (Goodfellow et al. 2016).

However, passing data to an untrained ANN will not yield satisfactory results. To train an artificial neural network we must use a procedure called *backpropagation*. Backpropagation is very similar to the process of learning by trial and error, and requires pre-labeled data. The name derives from the idea that, once the network is confronted with its error in the prediction, that new knowledge needs to be propagated back from the outcome layer of neurons to the hidden layers. In this process we first need to ask our ANN to perform a "prediction" using the feed-forward process. This is not yet a formal prediction, and is part of the training process.

The goal of this training prediction is to contrast the default values being output by our ANN with the actual real output values, such as classes. In this training prediction, all the matrices of weights are initialised with random numbers, so the outcome values are going to be very random and not related with the real ones. If the values are correct, nothing happens, but if a value is wrong, a difference between the prediction value and the actual value is computed. This difference (which is called error) is calculated by the *loss function*. This difference is then used to adjust the weights of the ANN in order to nudge them to the right configuration of weights to perform the best prediction possible, starting from the outcome layer all the way back to the first layer of the ANN.

The process of adjusting the weights is performed through small corrections every time the ANN is confronted with a new training data point. This process is called *gradient descent* (Ruder 2016), and its goal is minimising the output of the loss function (e.g., the amount of error in the predictions), by making small incremental changes in the weights of the ANN, so the outcomes of the loss function would gradually *descent* in magnitude after each iteration of the training process (e.g., every time the ANN is presented with a new training example) (Goodfellow et al. 2016). The process of calculating the gradient for the output layer L to a previous layer $L-1$ of just one weight w, assuming y to be the real value we are trying to approximate, can be understood as the following differentiation problem:

$$\begin{aligned} z^{(L)} &= w^{(L)} \times a^{(L-1)}; \\ a^{(L)} &= \sigma\left(\boldsymbol{z}^{(L)}\right); \\ C &= (a^{(L)} - y)^2; \\ \frac{\partial C}{\partial w^{(L)}} &= \frac{\partial C}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial w^{(L)}} = 2\left(a^{(L)} - y\right) \sigma'\left(z^{(L)}\right) a^{(L-1)} \end{aligned} \tag{4.18}$$

Equation 4.18 describes the process of taking the partial derivative of the loss function C with respect to the weight w of the current layer L. Since w is not a term in C, we start by finding the partial derivative of z with respect to $w = a^{(L-1)}$, then the partial derivative of a with respect of $z = \sigma'\left(z^{(L)}\right)$, the partial derivative of C with respect $a = 2\left(a^{(L)} - y\right)$, to finally multiply the three to obtain the value of $\frac{\partial C}{\partial w^{(L)}}$. The value of this partial derivative tells us how much and in which direction the weight should change in order to minimise the loss function.

To obtain the new value of the weight in the current layer L, and complete the process of *gradient descent*, the value of $\frac{\partial C}{\partial w^{(L)}}$ is then multiplied by a *learning rate*, which modulates the amount of change of the weights at each iterations, and is subtracted from the current weights as follows:

$$w^{(L)}_{new} = w^{(L)}_{current} - \text{learning rate} \times \frac{\partial C}{\partial w^{(L)}} \tag{4.19}$$

It is interesting to note in equation 4.18 that the partial derivative of the loss function with respect to the weight in current layer $w^{(L)}$ is directly dependent on the activation of the previous layer $a^{(L-1)}$, and hence it calculates how much the weights of $a^{(L-1)}$ affect the loss function and how much they should change to minimise the loss function. This is the process of *backpropagation* for one single weight. This process has to be repeated for all the weights, for all the layers in an ANN in order to properly adjust the whole system so the loss function can be minimised.

A fully working example of a simple multi-layer perceptron written in Python can be found in appendix A to illustrate in detail its inner workings.

Using the basic principles described above, ANN can be used to create much more complex classifiers that can adapt their weight matrices to incredible nuanced classifications problems. This makes them ideal for natural language processing classification, in which subtle changes in the configuration of words can change the meaning of a sentence.

Transformers or "Attention is all you need"

Vaswani et al. (2017)'s paper titled *Attention is all you need* came to revolutionise again the field of AI and deep learning by introducing a new type of neural network architecture capable of outperforming most of the previous ones in many of the most demanding NLP tasks: *The Transformers*. Its architecture follows the idea of having an encoder-decoder in which an entire sequence of text is encoded into a vector representation (Cho et al. 2014), but adding an important component: an attention mechanism.

Recurrent Neural Networks (Cho et al. 2014) are a particular kind of artificial neural networks designed to handle better sequential data. Sequential data are just data that comes and makes sense in a sequence. A prime example is text, which makes sense only as a sequence of characters or sequences of words. In order to make an artificial neural network understand the sequential

nature of the data being passed through it, a recurrent mechanism, in which the output of the network is fed again into it was proposed (Cho et al. 2014). This allows the network to have a memory of the previous information in the sequence, and helps the network to predict much better the next elements in an incomplete sequence. However, RNNs suffer from a problem known as *vanishing/exploding gradient*. This problem made RNNs fail at tasks, in which memory of long past data input was needed to predict the final output (e.g., a paragraph in which the subject is mentioned by name only at the beginning, but later only by its pronouns). Long Short-Term Memory (LSTM) Recurrent Neural Networks (Hochreiter and Schmidhuber 1996) tried to solve this problem by increasing the "memory" of the networks, applying *gates* to it. The gates in an LSTM-RNN are computing units very similar to regular neurons, but with connections to specific sections of the architecture, and thus holding information relevant only for that step of the process. For example, LSTM have *forget gates* which allows them to reset the information held in their memory units. These gates learn the right time when to activate from the data passed through the network in the training process. However, even with the great success of LSTMs, there were some NLP tasks in which the dependencies of the output were so far behind that not even LSTMs were enough. Then the paper of Vaswani et al. (2017) came out, proposing to replace the recurrent layers of LSTMs or RNNs with an attention mechanism. This attention mechanism is able to look into the sequence input and calculate an *attention score* for all the words/tokens (see Figure 4.7). This attention score is a representation of the strength of the relation of the words/tokens in the corpus used in the training process, and it is similar to a correlation matrix between words/tokens.

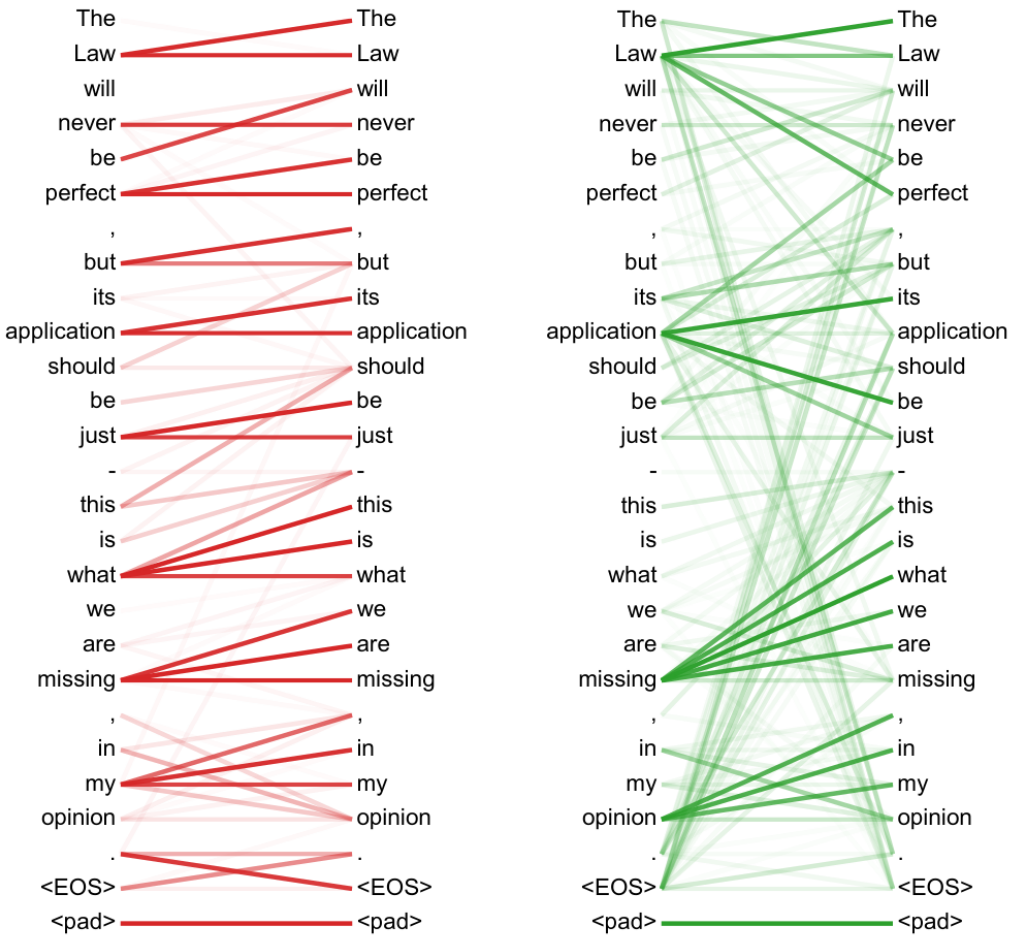

Figure 4.7: Two (green and red) different self-attention heads from a transformer model. The strength of the relationship between the words is represented by the opacity of the lines. Each head learned to attend to different parts of the sentence. While the green head seems to be leaning long-term dependencies between words, the red head seems to be focused on short term dependencies. From Vaswani et al. (2017)

The transformers attention mechanism allows the network to reference very long distance dependencies in the sequence, allowing them to learn the full structure of sentences, regardless of how long the input can be. This is achieved by consuming the whole sequence in parallel (instead of sequentially like LSTM-RNNs) and storing the contextual information in the attention heads. This contextual information is used later, for example, to make predictions of the next word in a sentence, by forcing the network to look at the attention scores of the words/tokens that has been seen in sentences at the left of (previous to) the tokens that needs to be predicted. If we were to use the example of attention scores presented in Figure 4.7, and we would like to predict the word "missing" using the attention scores of the green attention head, we would conclude that the most likely word is in fact "missing" because the five immediate words seen previously at the left of it ("this," "is," "what," "we," "are"), all have strong attention scores

towards it.

This excellent attention mechanism is very useful in the field of machine translation and speech recognition, but also allows for the development of models seeking comprehensive learning of the structure of a whole language. The Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018) is one such big language model based on transformers. BERT was originally pre-trained by Google to understand English using Wikipedia English articles and BookCorpus dataset, and then expanded to many other languages. This pre-trained model can be later fine-tuned using labelled data to solve many NLPs problems, including classification tasks. Since the model already knows the relationship of the words in the target language, the fine-tuning process is focused on teaching the model only the specific task, using regular backpropagation. This allows for the use of much less training data to achieve state-of-the-art results. In this project, I will test the performance of BERT (base, uncased version), DistilBERT (base, uncased version) and RoBERTa (base version), the three most popular BERT-based, general-purpose large language models according to Huggingface[5], in order to determine the most suitable one to be applied in the task of emotion classification. DistilBERT is described by its authors as a smaller, faster, cheaper and lighter version of BERT (Sanh et al. 2019). Sanh et al. (2019) claims that they have created a model of 60% of the size of BERT retaining 97% of its capabilities. DistilBERT was trained using the same corpus as BERT: a concatenation of English Wikipedia and Toronto Book Corpus. The "distillation" process of DiltilBERT involves using BERT as a teacher for DistilBERT in its training phase. Effectively, DistilBERT "learned" to copy BERT behaviour as well as it could by using only a fraction of the parameters of its teacher model. Given this, it is expected that both models behave very similarly. RoBERTa (Robustly Optimised BERT Approach, Liu, Ott, et al. (2019)), on the other hand, uses the same architecture as BERT (same number of parameters as $BERT_{large}$), but it takes a robust approach to choose the hyper-parameters of the training process as well as adding significantly more training data (160 GB of raw text in RoBERTa vs 16 GB in BERT). Liu, Ott, et al. (2019) summarises the modification to the BERT training process that led to RoBERTa as: " (1) training the model longer, with bigger batches, over more data; (2) removing the next sentence prediction objective; (3) training on longer sequences; and (4) dynamically changing the masking pattern[6] applied to the training data. We also collect a large new dataset (CC-

[5]Huggingface is website and Python library to download and upload pre-trained and fined-tuned large language models such BERT, RoBERTa and DistilBERT. Its URL is https://huggingface.co/

[6]Masking is the process of hiding words from a sentence, but keeping its position index, effectively communi-

NEWS) of comparable size to other privately used datasets, to better control for training set size effects."

The modification in the training hyper-parameters and training data used in RoBERTa are expected to produce differences in performance between BERT and DistilBERT. However, it remains to be seen if these differences in performance are significant enough to justify using it over lighter, faster alternatives.

Fine-tuning

As it was mentioned above, transformers-based models undergo a pre-training stage. In this pre-training stage the models are trained in "language understanding" on one or multiple languages, by performing a Masked-language Modelling (MLM) task. In this task, a large quantity of sentences are passed through the model, but in some of them a random word or words in each sentence are "masked" away, meaning that they are hidden for the model, but the model still knows that a word is missing. The model then needs to guess/predict which word is the correct one, and then compare it with the original sentence to obtain an error score. The transformer model is able to guess correctly by "reading" a lot of sentences and calculating the probabilities that a word comes before or after another. Thanks to this pre-training task, transformers models are able to learn the structure of a language.

However, MLM is not very useful beyond the pre-training stage. In order to make the transformers models useful for other NLP tasks, a **fine-tuning** stage is needed. In the fine-tuning stage, we usually use labelled data to further teach the already pre-trained model to perform a more specific task in the language it was trained on. In the case of this thesis, the fine-tuning task is to detect emotions in sentences (tweets). To do this, we follow the usual workflow of supervised machine learning training, in which we provide the model with labelled examples in a training step, and then we compute how good the predictions are in a testing step. Arguably, this could be done with a totally untrained transformer model, or any other deep learning model architecture. However, in those cases, we would need millions and millions of labelled sentences. The pre-training step allow us to reduce significantly the amount of data needed to train a model in a specific task to the order of only a few thousands of labelled sentences, because the model already "knows" the language, and can more easily generalise the correct

cating the model that the masked word is missing, and hence it must be inferred.

label (e.g., emotion) of similar sentences.

More formally, fine-tuning is an example of "transfer learning" (see Pan and Yang (2009)). In machine learning, transfer learning is the area of research in which knowledge encoded in a model, trained in a specific domain/problem, is used to solve a problem in a different but related domain/problem. For example, a model trained on recognizing cats in images could be applied to recognise dogs in images. Transfer learning will have greater success if the original task is more general than the final task. In the case of this thesis, the task of Masked-Language modelling is significantly more general than emotion detection, and thus the chances of successfully transferring the learning from the original task to the final task are very high.

Transformers-based classifiers Testing Setup

From transformers-based classifiers, and as mentioned in Chapter 4.2, I'm going to test three versions. For the English language dataset those versions are: BERT base uncased, DistilBERT base uncased, and RoBERTa base. For the Spanish version dataset, the versions are: BERT base multilingual cased, DistilBERT base multilingual cased and XLM RoBERTa base. The "base" suffix implies that these versions are the general purpose ones of each of the models and no fine-tuning has been applied to them. The "uncased" suffix denotes that the model takes "uncased" input and was trained with lowercase words. The "cased" suffix denotes the models were trained on sentences as they were written, with uppercase and lowercase letters. All models were fine-tuned with a 80:20 proportion for training and testing datasets. The fine-tuning session for each model consisted of 10 epochs (an epoch being a complete pass over the training data). After each epoch the state of each model was saved on a "checkpoint" file, and precision recall and F1 score were calculated, as well as the loss values for the testing dataset. For detailed information about the fine-tuning process of transformers based models see appendix Section B.1.

4.3 Data augmentation

With the increasing use of supervised machine learning models, the necessity for more training data increases as well. The digitisation of communication has helped to increase the available data, but the task of collecting, labelling and curating data for training purposes is still a daunting one. Moreover, the increasingly specific models face the challenge of hard-to-find

data points, leading to increasingly unbalanced datasets. As an answers to these problems, the concept of *data augmentation* was developed. Originally meant as a way of safely oversampling minority classes in very unbalanced datasets, the main idea is to generate new synthetic data by introducing some kind of variation on the original data. This method proved to be very popular in image recognition tasks (Shorten and Khoshgoftaar 2019) in which the introduction of small variations to the images allowed the researchers to obtain better generalisation results. However although popular in visual tasks and speech recognition, data augmentation is less popular in text-based natural language processing problems (Shorten, Khoshgoftaar, and Furht 2021; Wei and Zou 2019; Liu, Wang, et al. 2020). The main problem is that the information density of the analysis units in natural language processes is different from the amount of information carried by a pixel in an image, or a signal in an audio feed. Pixel level perturbations are unlikely to alter significantly the recognisable shapes in an image, and a similar thing can be said about small perturbations in sound. However, perturbations in words are much more likely to alter the meaning of a whole sentence. Data augmentation in text must take into account the semantic content of the sentence, and how this can vary when swapping words around (Shorten, Khoshgoftaar, and Furht 2021). Several methods have been proposed to tackle this challenge, including the use of synonyms, word2vec and rule-based replacements. However, those solution can only provide word-level replacements.

Recent developments in language modelling have allowed researchers to use Transformers-based translations models to create a pipeline of *backtranslation* which introduces context-aware, semantically stable variations of sentences (Beddiar et al. 2021). The methodology proposed by Beddiar et al. (2021) allows to create sentence variations that retain the same meaning of the original, without the risk of introducing disrupting variations (e.g., out of place words, random words, grammatical error, etc.). The newly generated sentence can then be added to the training dataset as a synthetic, but valid, new sample. Figure 4.8 illustrates the process.

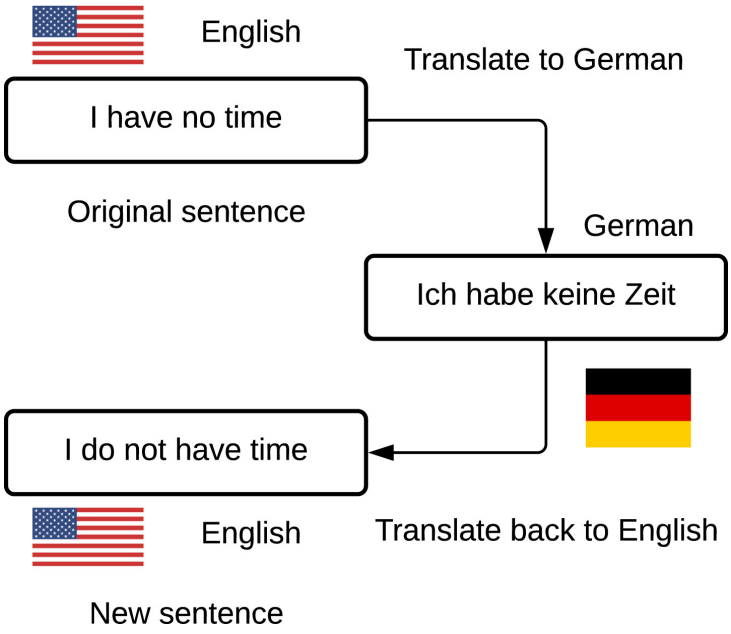

Figure 4.8: Backtranslation pipeline described by Beddiar et al. (2021).

In this thesis I will be using the methodology proposed by Beddiar et al. (2021), in which a Transformers-based translation model is used to generate new, semantic invariant, sentences from the original data. However, unlike Beddiar et al. (2021) I will use a pipeline including four languages in 2 batches. The first batch will use closely related languages to Spanish and English, namely German and French. The second batch will use two distant languages, namely Chinese and Arabic for the English dataset; and Russian and Arabic for the Spanish dataset. The discrepancy with respect to the distant languages was due to unavailability of Chinese-Spanish-Chinese translation models.

The main motivation behind the inclusion of more than one language in the backtranslation process is to smooth out possible systematic biases in the augmentation process that the use of only one translation model could introduce. This also has the beneficial side effect of introducing greater variability in the outcomes (e.g., the synthetic tweets).

In order to measure the similarity between original and backtranslated datasets I will use normalised Levenshtein edit distance (for detailed implementation see B.2), effectively comparing the number of edits (character swaps, deletions or additions) necessary to transform the translated version into an exact copy of the original sentence.

4.4 Knowledge Discovery

In the task of "aspect identification" (Group-based Motivator, Collective Actions, Ingroup and Outgroup; see Section 2.5 for more details), the use of supervised machine learning techniques is not possible. The four socio-psychological aspects of a social movement I am set to detect here are composed of different entities (e.g., mentions of authorities, actions, groups, etc.) in each of the social movements under study. No previous work labelling the presence of such specific aspects in social movements tweets exists. That is, there are no labelled data for training and fine-tuning, hence no supervised machine learning method can be used to detect such aspects. The use of techniques such as parts-of-speech detection or named entity recognition would not be much of a help either since each aspect is composed of a specific combination of the latter, which is, again, specific for each social movement.

Given the above mentioned limitations, the strategy used in this thesis to detect mentions of specific aspects (Group-based Motivator, Collective Actions, Ingroup and Outgroup) inside tweets will be to identify the keywords representing these aspects in the corpus of tweets of each social movement. This process will be carried out by a set of knowledge discovery techniques combined with previous knowledge about each of the social movements.

Three main exploratory knowledge discovery analyses were performed in order to determine the appropriate set of keywords describing the four aspects of the social movements in study: word frequencies, bigram networks and topic modelling with Latent Dirichlet Allocation (LDA). Word frequencies is a straightforward analysis that consists in counting the most common words in a corpus. The other two techniques, however, are more complex, and we will benefit from a review of their inner workings in the following section.

4.4.1 Unsupervised Machine Learning methods

So far, I have reviewed supervised machine learning methods. Those methods are great for when the task can be addressed with labelled data, as it is the case with emotion recognition/-classification. However, there are some tasks that cannot be addressed this way and require an exploratory approach. Here is when Unsupervised Machine Learning is useful.

Unsupervised learning algorithms requires no labelled data, since their objective is usually to extract information that is hidden even for the annotator or expert. A classic example of the

problems that unsupervised learning tries to address is clustering. Clustering involves finding groups of observations in the data that are similar to each other and build a group or cluster. From these observations we only know their features but we do not know the group they belong to nor how many groups there are in the data. Going back to our flower example, in this case we would know only their sepals and petals width, and the task would be to identify if they form groups and how many groups best describes the data. In this case there is no training or testing datasets, since all data are analysed at once, and, although there are some metrics that help evaluating the fit of the algorithms to the data, their final performance assessment depends highly on the researcher and how informative the results are for the research.

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is an unsupervised machine learning technique commonly used in natural language processing as a method of knowledge discovery. Its goal is to find unobserved variables that might explain the variability of a set of observations. In the case of natural language processing, these observations are usually words that are usually organised inside some kind of document (e.g., letters, chapters, tweets). The algorithm assumes each document is composed of a random mixture of latent (unobserved and unknown) topics. LDA assigns each word in the corpus (all the documents) its own probability to belong to each of the topics. A topic is the collections of words which appear more often together across all the documents.

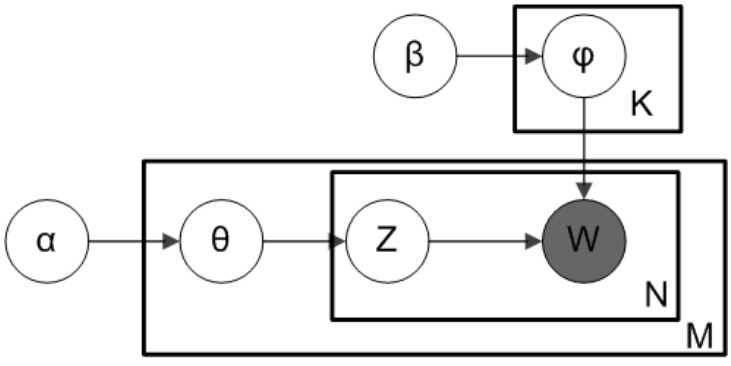

Figure 4.9: Plate notation for LDA with Dirichlet-distributed topic-word distributions. Adapted from Blei et al. (2003)

Figure 4.9 represents in graphically the most common form of LDA. The circles represent variables, while the rectangles represent repeated entities, which in this case are words or documents. W, shaded in grey, represents the only observable, which is a vector containing

all the words of the corpus. M denotes the number of documents and N denotes the number of words in a given document. Z is the topic assigned to each word in a given document. This is done based on the topic distribution Θ, estimated for each document in M. α is a hyperparameter (specified by the user) controlling the per-document topics distribution. Small values of α (less than 1) are preferred in order to have few topics per document. β, the Dirichlet prior, is a parameter that controls the distribution of words per topic. Very small values (e.g., 0.001) are preferred in order to obtain topics with fewer words. Finally, ϕ is the word distribution per topic K (Blei et al. 2003). Effectively, LDA returns the probability of each word to be part of each topic, and the probability of each document to belong to each of the extracted topics. With this information it is possible to extract which words are more representative of each topic and also to classify each document belonging to a respective topic.

Blei et al. (2003) proposed that the performance of the process can be evaluated by a perplexity score, described in equation 4.20. A lower perplexity value indicates better generalisation performance.

$$per(D_{test}) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right) \tag{4.20}$$

In equation 4.20, M is the number of d documents in the corpus, $p(w_d)$ is the assigned probability of each word in its corresponding d document, and N_d is the number of words per document d. The perplexity is, in essence, a measure of how well a probability model predicts the distribution of a new set of data, based on the probabilities inferred in the training data. In this case, a higher perplexity would indicate a bad fit to the data, and thus a bad description of the sampled corpus.

LDA is an excellent technique for knowledge discovery on large text corpus, and I will use it to explore the common targets of the emotional response inside each social movement (see Chapter 3).

N-grams, Bigrams and Bigram networks

N-grams are a very simple notion introduced by Shannon (1948) work in information theory originally thought as a way of calculating the likelihood of the next letter based on the previous

one. Bigrams are an extension of this used to calculate the likelihood of a word given a previous one as follows:

$$P(W_n|W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})} \tag{4.21}$$

Where the probability $P()$ of a word W_n given the preceding words W_{n-1} is equal to the co-occurrence of the two words $P(W_{n-1}, W_n)$ divided by the probability of occurrence of the preceding word $P(W_{n-1})$. This probability can be used to detect which co-occurrences are common in a given corpus of text, which in turn can be used to inspect the most common topics of the corpus.

Bigrams can also be understood as a nominative list, where words nominate each other's relations. This nominative list can be easily turned into nominative networks and then visualised as a network mapping the relations of common words in the corpus. In this thesis I will use this technique to help me find the common topics in the corpus of tweets. By visualising the words that frequently appear together I can infer the topics of the conversations users had in Twitter.

4.4.2 Knowledge Discovery Set-Up

Not all of the tweet data per movement (see Table 4.1) was used in knowledge discovery analyses. In order to reduce the computational burden while keeping representativeness, a random sample of 100,000 tweets per movement was extracted for the purpose of conducting this analysis. These samples were submitted to a process of text cleaning in which common stop-words[7], punctuation, user tags, URLs and emojis were removed. Some extremely common words in the datasets were also removed to perform Words Frequencies and LDA, since in preliminary analysis, they were present in all of the LDA topics. In the case of of the Chilean 2019 Social Movement tweets dataset, the terms removed were: "chile," "carabineros." In the case of the Hong Kong 2019 Social Movements the terms removed were: "hongkong," "hong," "kong." Finally, in the case of Fridays for Future, the terms removed were: "climatestrike," "fridaysforfuture." In the case of bigram networks, the frequency of words can be used as a parameter of the node, hence for this analysis, no extremely common words were removed.

Finally, topic modelling using Latent Dirichlet Allocation requires deciding the number of topics

[7]Stop-words are words such as "the," "a," "to," etc., used as building blocks of the syntax of a language but that cannot carry much information.

to be extracted from the corpus. Although this decision must be made taking into account the context knowledge the researcher has on the issue and the objective of the analysis, metrics on the optimal amount of topics for a given corpus can be calculated to further inform the decision. In this thesis I will use the perplexity score of the LDA topic solution (see Equation 4.20 for details) as a performance metric to decide the number of topics to extract in each case. I will calculate the perplexity score of the LDA topic solutions from two to ten topics. I stopped at ten mainly because of the interpretability of solutions beyond ten topics. Given that my goal is to find topics that relate to four aspects, I considered that ten topics is a good upper limit for the exploration.

4.5 Time-series analysis techniques

So far, I have described the methods that will help me analyse the text of the tweets, and transform that text information into numerical information. After that, I will aggregate that numerical information by units of time (e.g., days), and analyse it as records of information over time. This data format is known as time-series.

Time-series is the name given to a sequence of data points indexed in time order. It is common, although not necessary, that the sequence is taken in equally spaced periods of time, forming a discrete-time dataset. Common examples of this kind of data are: GDP over time, stock values, number of cases of a disease per day, etc. In my case, an example metric could be the number of "angry" tweets per day or the number of "sad" tweets per day.

4.5.1 Auto-Regressive models

An important characteristic of time-series data are that independence between data points cannot be assumed. More precisely, there is a reasonable possibility that the current data point is dependent on the previous one. To correct for this effects, techniques modelling the behaviour of time-series incorporate an auto-regressive component (see Cryer (1986), p. 66), controlling for the effects of the value of the previous data points over the current one as follows:

$$X_t = c + \sum_{i-1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{4.22}$$

In the above equation 4.22, the value of the current data point X_t is determined by the linear

model with a constant c, degree p, parameters $\varphi_i \dots \varphi_p$ and error ε_t.

4.5.2 Vector Auto-Regressive models

Auto-regressive models are great when only one time-series needs to be analysed, but fall short when multiple time-series are included into the analysis. Vector Auto-Regressive (VAR) models (Lütkepohl 2013) come to solve this issue by incorporating the new time-series in a generalisation of the auto-regressive model. A VAR model with p lags (number of back steps) of k variables, is written as:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t \tag{4.23}$$

Where c is now a vector of k constants serving as the intercepts of the models. A_i is a time-invariant $(k \times k)$ matrix and ε_t is a vector of k error terms.

The main advantage of VAR models is that they can be used to calculate the influence of several time-series on another time-series, and include the estimation of the *Granger causality* (Granger 1969). *Granger causality* allows to determine statistically whether a time-series can significantly predict another, and is defined as:

$$X_t = \sum_{\tau-1}^{L} A_\tau X(t-\tau) + \varepsilon_t \tag{4.24}$$

Where L is the number of time lags τ, A_τ is a matrix of variables for every τ. A time series X_j is Granger caused if at least one of the elements of the matrix A_τ is significantly larger than zero. Since Granger has an approximate F distribution, an F statistical test can be used to obtain the p value associated with the Granger score.

The Granger causality test does not evaluate "true causality" and its name is a misnomer. The test itself can only provide information about how well a time-series forecasts another. In this thesis I will use Granger Causality to assess how well the time-series containing the daily measures of the emotions of each social movement, forecast the time-series containing the records of events (violent or non-violent) of the corresponding social movement.

4.5.3 Vector Auto-Regressive Models Setup

The time-series data of each social movement will be analysed using Vector Auto-regressive (VAR) models (see Section 4.5.2) in order to determine any possible causal relationships between the emotions and violent and non-violent events. Before running the VAR analysis, the count data of each emotion was transformed using a log function in order to reduce the range of the values which could be between zero and hundreds of thousands. This transformation helps to improve the fit of the model and prevent integer overflow[8] in the computations.

Additionally, VAR models require all the variables to be stationary, meaning that no time dependent trends should be present in the variables in order to avoid spurious regressions. All variables, including the already log transformed emotions time-series, were tested for stationarity using Augmented Dickey-Fuller unit root test(Dickey and Fuller 1979). Where stationarity was not achieved[9] further "differencing"[10] transformation was applied to the variable in order to achieve stationarity. All variables, which were transformed using differencing, are denoted in the Tables with the prefix ∇.

VAR models are required to provide a 'lag' or 'order' parameter for the equations. These values can be translated as how many periods (days in this case) the model is going to look backwards for auto-regressive effects. While there are methods for estimating the optimal amount of lags in a VAR model, in this thesis I will follow a theoretically driven criterion to select the amount of lags, setting them at exactly 1 day back $(t-1)$. This is because I do not have any theoretical reason to believe longer lags could be meaningful. Emotions, as discussed in Section 2.3.1, are fleeting mental states, evolved to motivate immediate action in humans. In order to keep an emotion active for long periods of time, memories and cognitive elaborations should be present, transforming that mental state into an 'attitude'. For this reason, the VAR models will restrict the effects of emotions on the collective actions (and vice-versa) to one day. Nevertheless, analysis of the optimal amount of lags will be provided for every VAR model, in order to provide supporting evidence for this theoretical decision.

Finally, given a lag of 1, and a number of k variables y, the models will have the following form:

[8] Integer overflow is a type of low level computing error which occurs when the processor is presented with integers bigger than what is able to represent based on its architecture. For example, in a 32-bit processor the largest unsigned integer possible is 4,294,967,295.

[9] In Augmented Dickey-Fuller (ADF) test failure to reject the null hypothesis leads to the assumption the tested time-series are non-stationary.

[10] A differencing transformation takes the difference of the current value and the previous one in a time series.

$$\begin{bmatrix} y_{1t} \\ \vdots \\ y_{Kt} \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix} + \begin{bmatrix} \beta_{11} & \dots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{K1} & \dots & \beta_{KK} \end{bmatrix} \begin{bmatrix} y_{1(t-1)} \\ \vdots \\ y_{K(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Kt} \end{bmatrix} \tag{4.25}$$

Where the vector of variables $[y_{1t} \dots y_{kt}]$ represents the values of the variables at the current time t with no lags (e.g., In this case, the values of violent and non-violent reports and the values of the emotions, per day); $[c_1 \dots c_k]$ represents the values of the intercept for all the k variables; $\begin{smallmatrix} \beta_{11} & \dots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{k1} & \dots & \beta_{kk} \end{smallmatrix}$ represents the values of the estimates for the models (or betas); $[y_{1(t-1)} \dots y_{k(t-)}]$ represents the values of the variables lagged one period; and finally $[\varepsilon_{1t} \dots \varepsilon_{kt}]$ represents a vector of normally distributed errors for each of the k variables. For each social movement, five VAR models were estimated: A general model using the emotional trends detected in all the Twitter filtered data as predictors of Violent and Non-violent collective actions; and four models for the effects of the emotional trends in relation to each of the four socio-psychological aspects. The logic behind this is to first determine if the general pattern of emotions in Twitter predicts violent or non-violent collective actions and then to determine if any of the emotions in relation to each of the socio-psychological aspects can predict, separately, the occurrence of violent or non-violent collective actions. It must be noted that VAR models do not have a unique dependent variable, since the process makes estimates for each time series variable as an outcome of the lagged and unlagged values of all the time series variables, including itself. This results in a matrix of estimates for each model, similar to a correlation matrix. Nevertheless, for simplicity, in this thesis I will assume that the main dependent variables of the model are the reports of violent or non-violent collective actions, although effects of collective actions on emotions will also be noted and discussed.

Chapter 5

Results

As stated in Chapter 1 in this Chapter I will present the results that will help me identify the emotions of anger, fear, sadness and joy in the tweets associated with the three social movements presented in Chapter 3, and how these expressed emotions evolved. Using time-series analysis I will establish the relation of these four emotions with real-life events and collective actions performed by each of these social movements.

The chapter is divided in the following sections:

1. Methodological Approach Selection: In this section I will present and discuss the results of the process of selecting the best classification models for emotion detection.

2. Data Augmentation: In this section I will provide the data augmentation results. This procedure aims to further improve the accuracy of the selected emotion classification models.

3. Aspect Identification: In this section, using the knowledge discovery techniques, topics modelling and bigram networks, I will identify the keywords related to the fours aspects (Group-based motivator, Collective Actions, Ingroup, Outgroup) present in the tweets corpus of each of the social movements studied.

4. Emotions in social movements: In this section, the classification results of the selected model are presented for each social movement. Here, the evolution and patterns of emotions are presented.

5. Finally, in this section I will discuss the results and their theoretical and methodological

implications. In particular, I aim to discuss how these results relate to the political psychology models presented in Section 2.4.

5.1 Methodological Approach Selection

As outlined in Section 4.2, an important part of my research deals with the assessment of text classification techniques and how appropriate they are for social science research. In this project I relied on methodologies to infer the values of my key variables, namely, the emotion categories contained in a tweet, from text data. Such methods provide only proxy measures compared to more direct methods, such as direct observation (e.g., face-to-face interviews, physiological measures), or even self-reports of emotional states. This weaker link affects directly the epistemological validity of my measures. Given this, I must make sure that the inference process used in the emotion detection has the highest degree of accuracy current methodologies allow to achieve, in order to have confidence in results, when analysing the relationships of the tweets' emotions and real-life protest events. In order to achieve such high validity and accuracy I conducted a comparative study on the most popular techniques for emotion detection from text data and measured their accuracy against human annotated gold standard datasets (see 4.2). In the following sections I will present the results of this study. Given that my data were multilingual (English and Spanish), two parallel studies were conducted to identify the best classification technique for both English and Spanish. For that reason, the model selection results are being presented by language, with parallel subsections for English and Spanish.

5.1.1 Classification Methods Performance

In this section I will provide a review of the performance of the classification methods reviewed in Chapter 4, namely, Lexicon-based methods, Support Vector Machines methods (SVM), Naive Bayes methods (NB), and Transformers-based methods. Each method will be evaluated using precision, recall and F1 (raw, macro and weighted, see 4.2.3 for details) scores. Structure wise, I will first provide an overview of the results to later dive into the details for each method.

As can be seen in Table 5.1, Transformers-based classifiers, and specifically BERT, dominated the classification scores for the English dataset. It is worth nothing, nevertheless, that Linear Support Vector Machines and Complement Naive Bayes still perform admirably well under the conditions of this test. Under simple conditions and when speed efficiency is more important

than accuracy in the prediction, SVMs and Naive Bayes are still strong competitors worthy of consideration. On the other hand, Lexicon-based methods performed worst and their use is not recommended for NLP classification tasks.

Classifier	Precision	Recall	F_1Score	F_1Macro	$F_1Weighted$
NCR Lexicon	0.15	0.35	0.12	0.14	0.13
SVM Linear	0.88	0.90	0.91	0.89	0.91
SVM RBF	0.84	0.89	0.88	0.86	0.88
Multinomial NB	0.54	0.81	0.68	0.54	0.74
Complement NB	0.86	0.89	0.88	0.87	0.89
Bernoulli NB	0.65	0.79	0.75	0.67	0.78
BERT	**0.93**	**0.92**	**0.94**	**0.93**	**0.94**
DistilBERT	0.92	**0.92**	0.93	0.92	0.93
RoBERTa	0.91	0.91	0.92	0.91	0.92

Table 5.1: Classifiers metrics for English language dataset. Higher scores indicate better performance. Highest values per column in bold.

For the Spanish dataset, the results were less decisive. Although in Table 5.2 we can see that XLM RoBERTa is the best model, the second best one was not any of the transformers-based variants but the linear support vector machines, which performed very well under the conditions of this test, reinforcing the idea that under constraints of efficiency SVM should be seriously considered.

Classifier	Precision	Recall	F_1Score	F_1Macro	$F_1Weighted$
SVM Linear	0.74	0.75	0.77	0.75	0.77
SVM RBF	0.72	0.74	0.75	0.73	0.75
Multinomial NB	0.61	0.72	0.68	0.62	0.71
Complement NB	0.73	0.74	0.75	0.73	0.76
Bernoulli NB	0.64	0.69	0.68	0.65	0.70
BERT	0.73	0.74	0.76	0.73	0.76
DistilBERT	0.73	0.72	0.75	0.73	0.75
XLM RoBERTa	**0.77**	**0.77**	**0.79**	**0.77**	**0.79**

Table 5.2: Classifiers metrics for Spanish language dataset. Higher scores indicate better performance. Highest values per column in bold. Note: NCR Lexicon was not available in Spanish.

In Table 5.1 I presented the results of the performance evaluation of nine emotion classification methods for English and in Table 5.2 the results for eight emotion classification methods for Spanish, given that the lexicon based method was only available in English. Given these results, I will now discuss the performance of each of the four main approaches, Lexicon-based, Naive Bayes, Support Vector Machine and Transformers-based models in greater detail.

5.1.2 NCR lexicon performance

First, let's take some time to analyse the case of the NCR lexicon and its very low performance. These results can be explained by a number of factors, but probably one of the most important ones is that it was constructed for a greater number of emotions (eight in total) than the ones used in this study (only four: anger, fear, joy, sadness). The test set-up was also rather strict for the NCR lexicon. Prediction of similar (although still erroneous) emotions was considered incorrect in the evaluation rules (see Section 4.2.4). This explains in part why its accuracy is so low. This means that in the hypothetical worst case scenario in which every algorithm made completely random guesses of the emotion of each of the testing tweets, the probability of the machine learning based methods of guessing right is 1 in 4 (per case), whereas in the case of the NCR lexicon, it was 1 in 8. This is something that could be corrected by adapting the lexicon itself, however, that solution is rather costly considering that there are already other, relatively simple, machine learning solutions that outperform lexicon solutions.

In addition to the above mentioned structural disadvantage, lexicon methods are notoriously static and rigid. Lexicon development requires considerably more resources than the annotations of tweets examples. Continuous adaptation of the lexicon to the ever evolving meaning of words is simply not feasible. In addition, its rigid state imposes constraints on the classification task that other techniques do not. For example, it is possible to imagine other scenarios, in which the target emotions are not present in the list of emotions of a lexicon, limiting the application of it. In this study, however, NCR faced a favourable scenario in which all the emotions to classify were present in the lexicon.

Finally, the NCR lexicon follows a rather primitive word-emotion structure. Each word/term has an emotion value associated with each of the emotions included in the lexicon. This means that it does not consider the position of the word within a sentence or modifying words such as "very."

Considering all of the above mentioned restrictions and the overall accuracy achieved by the NCR lexicon, this approach will not be used for obtaining results in the context of this study.

5.1.3 Naive Bayes performance

Naive Bayes classifiers are an extremely fast and computationally efficient family of classifiers. They can work efficiently with very high dimensional data (many features per case), which is a

characteristic that makes them very attractive for text classification. However, this comes with several disadvantages. A Naive Bayes Classifier is, unsurprisingly, very naive in its assumptions about how the data behaves. Notable, it assumes that there are no dependencies between any pair of features that can affect the classification (Zhang 2004). Specifically, the algorithm was proposed for data for which the assumption holds that dependencies between features are normally distributed and thus cancel each other out. (Zhang 2004).

In general, naive assumptions in algorithms are not a problem as long as the algorithm performs with acceptable accuracy. In the context of this study the Naive Bayes Classifier landed on third place in terms of overall accuracy. So the performance was good and yet, with sequential data, such as text, it is not correct to assume that features do not have inter-dependencies. Neither do these dependencies cancel each other out. Specifically, in text data, features (words) generally depend on features (words) following after. In addition to this general structure of the language, there are special cases like modifier words, word positions and long-distance dependencies (e.g., Nouns mentioned at the beginning of a sentence, but then referred only by its pronoun) that act as context, which, under the Naive Bayes assumptions, would be completely neglected. However, there are also important differences between different classifiers from the Naive Bayes Classifier Family (see Section 4.2.4 in Chapter 4). I have tested three different Naive Bayes Classifiers: the Complement Naive Bayes Classifier, the Bernoulli Naive Bayes Classifier and the Multinomial Naive Bayes Classifier.

The Complement Naive Bayes Classifier, in particular, greatly outperformed the Multinomial and Bernoulli classifier versions. This was partially expected. The Complement Naive Bayes Classifier was designed to correct the severe assumptions of the Multinomial Naive Bayes Classifier regarding the balance of training datasets (Rennie et al. 2003). However, its superiority over Bernoulli Naive Bayes was not a given. In fact I expected that the Bernoulli interpretation of word distributions would help with classifying short text such as tweets. A Bernoulli Naive Bayes classifier takes into account only the presence (Yes or No) of a word in a document, instead of its frequency (count). This should favour its performance in short text, in which repetitions of words are less common (Metsis et al. 2006; Singh et al. 2019; McCallum, Nigam, et al. 1998). However, while the Bernoulli Naive Bayes Classifier did perform better than the Multinomial Naive Bayes Classifier, it was nevertheless out-competed by the Complement Naive Bayes Classifier. I believe these results provide important insights about the task of emotion

classification in tweets, instead of providing information about the classifiers themselves. Complement Naive Bayes accounts for the imbalances in the training datasets, for the multinomial characteristics of the features (e.g., words frequencies), and for the repetitions of words inside the documents. Words repetitions do not provide a great amount of explicit new information to the readers, but they do provided implicit *emphasis*, which can be potentially important when classifying emotions in a text.

However, given that other methods performed better in the performance test, Naive Bayes classifiers will not be used in this work to detect emotions in tweets.

5.1.4 Support Vector Machines performance

In the results presented in Tables 5.1 and 5.2, Support Vector Machines classifiers performed admirably well despite their simplicity. Recalling what was discussed in Chapter 4, SVMs approach the problem of classification from a geometrical point of view, attempting to create boundaries in the feature space to separate the classes, instead of, for example, relying on probabilities like Naive Bayes classifiers do. This geometrical approximation to our classification problem appears to have given SMVs the upper hand. The results also appear to suggest that the emotion classification problem is, largely, linearly separable, given that the Linear SVM is the one showing the best performance among the two SMVs explored: SVM Linear and SVM Radial Basis Function (RBF). Let's unpack the meaning of this finding. As was mentioned in Section 4.2.4, Support Vector Machines take numerical features to create a feature space in which each data point is placed. In this case, the feature space is composed by the frequency count of each word per tweet. To better understand this, let's imagine a trivial example in which we have 20 tweets to classify each composed of only two words ("sad" and "happy") that are repeated several times. The feature space, formed by the frequency count of the words in these tweets, can be visualised as a plane, in which each axis represents the count of the words and each point is a tweet.

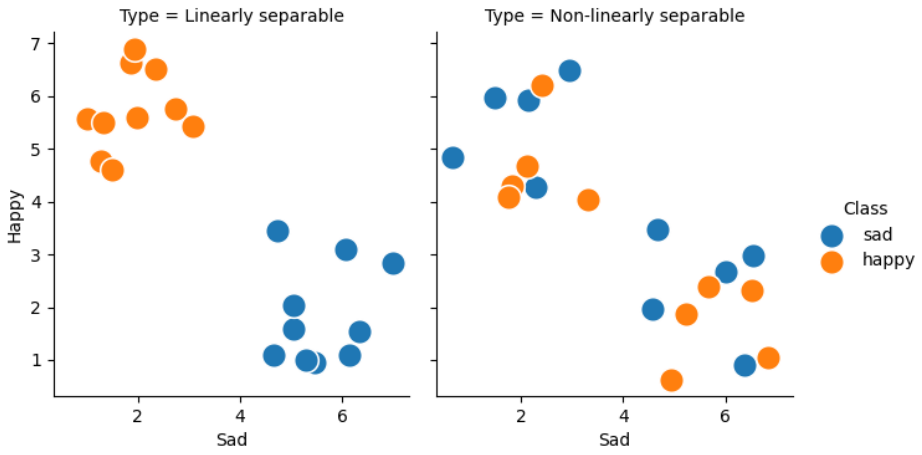

Figure 5.1: Example of (left) linearly separable and (right) non-linearly separable classifications problems using word frequencies features space. Note: Jittering was added to the data points for presentation purposes. Source: Personal collection.

Figure 5.1 shows the two possible relevant scenarios we might encounter in our trivial example. The plot on the left side of Figure 5.1 represents, in a very simplified way, the results obtained here with Linear SVMs. The amount of times a word is repeated in a document does not change the class membership (in this case, the emotional meaning) of that document. On the other hand, if the amount of repetitions of a word would have had an effect on the class membership of a document (in my case, an effect on the emotional meaning of the document) we would have observed a pattern similar to what the right side plot in Figure 5.1 presents. This finding is relevant because there are some plausible situations in which the emotional meaning of a sentence can change when a certain amount of repetition of a word is reached as is the case, for example, with sarcastic exaggerations or sarcastic hyperbole. The fact that the Linear SVM performed very well here does not mean that sarcastic exaggerations do not exist in the testing dataset, but rather that the human annotators, who took part in the construction of the testing and training datasets, weren't able to detect it if it was present, and that most likely, they interpreted repetitions (consecutive or nonconsecutive) as way to add emphasis to the message, similar to what the results of the Naive Bayes Classifiers suggests.

These findings highlight the need for good quality training and testing datasets, but also the importance of managing the expectations of what is doable with the amount of information the data are providing. Very short texts, like tweets, do not provide much contextual information. Usually, only the text of the tweet is provided for annotation, without any reference to the

conversation it was inscribed into, or the characteristics of the account that posted it. When annotators are faced with such a task they are very likely to fill the gaps in the contextual information with their own knowledge, which will introduce some noise in the labelling, and thus, will reduce the quality of the labelled data. In the case of the work presented in this thesis, I believe, however, that the expectations are well managed. Choosing a simpler taxonomy of four basic emotions instead of the more widely used taxonomies of six or eight basic emotions helps reduce the ambiguities in the classification process.

Despite the success of SVMs in the performance test, they will not be used in the final emotion detection task of this thesis, since Transformer-based methods proved to be more accurate and flexible, with an acceptable trade-off in training and computation time.

5.1.5 Transformers-based models performance

Transformers-based models were the ones performing best in terms of accuracy, being clear winners in the English classification task and also providing the best results with XLM RoBERTa for the Spanish classification task. This was expected. Transformers-based models, and in particular, BERT-based models have been dominating the scores in every NLP task in the last three years, thanks mainly to their ability to learn short and long distance dependencies between words. Now, compared to the closest competitor in the results presented in this project, the SVMs, Transformers-based models have the advantage of sophistication and complexity (millions of parameters to adapt), which gives us much more room for improvements compared to simpler classifiers. In other words, Transformers-based models could learn much more complex patterns if better and more varied fine-tuning data were given to them.

The significant differences in performance between the different Transformers-based approaches were not expected though. My explanation is that given the characteristics of the data I have for the fine tuning procedure (short text with little contextual information), these differences have likely more to do with the data used in the pre-training phase than with differences in the algorithm architecture. As it was stated in Section 4.2.4, Transformers-based models undergo a pre-training process, in which a large corpus of very varied text data are used to train the model to "understand" the underlying general structure (word relations) of a language. Pre-training two identical transformers-based models on different corpus, even of the same language, will yield slightly different results. Both BERT and DistilBERT were pre-trained using the same

corpus (and hence they have very similar results), whereas RoBERTa was trained on a different corpus, using a different configuration of hyper-parameters (see Section 4.2.4). In order to determine the best Transformers-based model to use, a more varied fine-tuning dataset must be used.

The differences in performance by language, on the other hand, have much more to do with the amount and variety of the data used in the fine-tuning stage. The Spanish training dataset was significantly smaller than the English, and significantly less varied, as it was only composed of tweets (whereas the English datasets also included short sentences from other sources). In order to improve the results, a larger more varied dataset should be used.

The above mentioned limitations are addressed in the next Section 5.2 in which an augmented fine-tuning dataset is used to test the accuracy of the BERT, DistilBERT and RoBERTa for English and Spanish.

Overall, given that Transformers-based models were the ones performing better in the emotions classification tasks, and given their potential for improvement, they will be the ones used further in this thesis. However, in order to determine which of the tested Transformers-based models should be chosen I will be consulting the results from the data augmentation study below.

5.2 Data Augmentation Study

5.2.1 Data augmentation using backtranslation

In Section 4.3 I introduced the concept of data augmentation, and outlined the challenges of data augmentation for text data in comparison with data augmentation for images and audio signals. To briefly recapitulate, data augmentation attempts to create artificial data from real data, by introducing small perturbations without affecting the overall features of the source data. In the case of this thesis, I will use backtranslation to new, slightly different, versions of the original sentences in the training data, effectively creating a synthetic version to add to the training data, and increase its size and variability.

Figures 5.2 and 5.3 presents the results of the backtranslation process, showing the distribution of normalised amounts of edits for each datasets, where 0 represents identical sentences between original and augmented text, and 1 represents complete swapping of characters between original and augmented text.

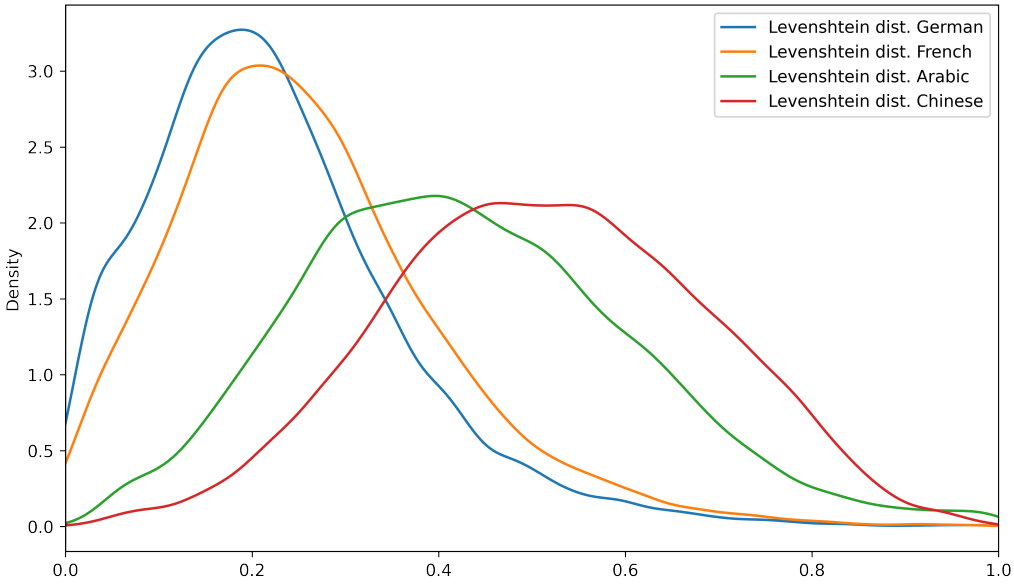

Figure 5.2: Distribution of the Levenshtein edit distance having as reference the original English text for the backtranslation in French (yellow), German (blue), Arabic (green) and Chinese (red). Source: Personal collection.

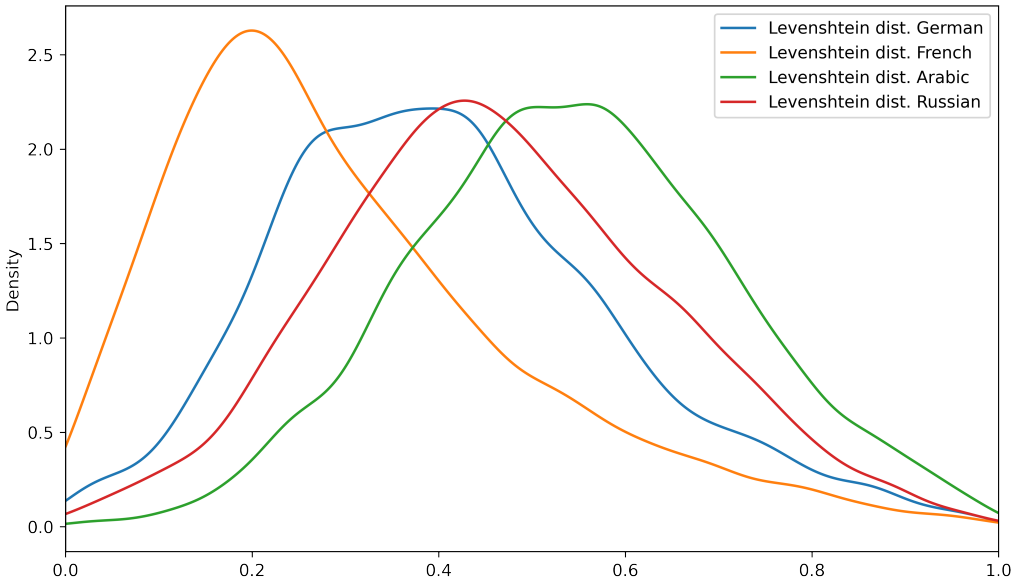

Figure 5.3: Distribution of the Levenshtein edit distance having as reference the original Spanish text for the backtranslation in French (yellow), German (blue), Arabic (green) and Russian (red). Source: Personal collection.

As can be seen in Figures 5.2 and 5.3, close languages (German and French) do produce less variant results than distant languages (Arabic, Chinese and Russian). In the case of the English dataset (see Figure 5.2), German and French backtranslations required around 20% of character edits and swapping to reach a sentence identical to the original, whereas for Arabic and Chinese that number increases to around 40% of character edits. In the case of the Spanish dataset the results follow a similar pattern, with German and French requiring around 20% and 30% of character edits respectively, and Arabic and Russian requiring 40% and 50% of character edits respectively.

From Figures 5.2 and 5.3 it can also be inferred that in some cases the backtraslation process produced identical results to the original sentence (edit distance equal to 0), causing the inclusion of duplicates in the data. To deal with this issue, those cases were deleted before constructing the augmented datasets.

In order to produce the augmented datasets, backtranslations from German and French were grouped together and added to the original to produce the *Near* augmented dataset. The same process was done for Arabic and Chinese (in the English case) and Arabic and Russian (for the Spanish case), producing the *Distant* augmented dataset. The separation between Near and Distant augmentations was maintained in order to evaluate the effect of backtranslation-induced variability on the performance of the classifiers to be trained on the augmented data.

Finally, and based on the performance results presented in Section 5.1.1 only transformer-based models were tested with augmented data. I assume that the performance of both English and Spanish classification Transformers-based models can benefit from a more varied, augmented training dataset.

5.2.2 Performance of Transformers-based Models with Augmented Data

In order to evaluate the performance of the Transformers-based models using augmented data, each of them was trained and evaluated with the *Distant* and *Near* version of the augmented data. The results in terms of precision, recall and overall accuracy (F1 scores) are presented in Tables 5.3, for the English case and 5.4 for the Spanish case.

Classifier	Precision	Recall	F_1Score	F_1Macro	$F_1Weighted$
BERT Distant	0.82	0.82	0.84	0.82	0.84
DistilBERT Distant	0.82	0.82	0.84	0.82	0.84
RoBERTa Distant	0.82	0.81	0.83	0.81	0.83
BERT Near	**0.93**	0.93	**0.94**	0.93	**0.94**
DistilBERT Near	**0.93**	0.92	**0.94**	0.93	**0.94**
RoBERTa Near	**0.93**	**0.94**	**0.94**	**0.94**	**0.94**

Table 5.3: Classifiers metrics for English language augmented datasets. Higher scores indicate better performance. Highest values per column in bold.

As can be seen in Table 5.3, for the case of the English dataset, all transformers-based models performed worse when trained with datasets augmented with backtranslations using distant languages. On the other hand, the performance of the models improved when using training dataset augmented with backtranslations from near languages. This suggests that some of the semantic information is lost in the process of backtranslation when the intermediate language is very distant from the target language.

Classifier	Precision	Recall	F_1Score	F_1Macro	$F_1Weighted$
BERT Distant	0.77	0.77	0.80	0.77	0.80
DistilBERT Distant	0.76	0.75	0.79	0.75	0.79
RoBERTa Distant	0.79	0.78	0.82	0.78	0.82
BERT Near	0.84	**0.85**	0.87	0.84	0.87
DistilBERT Near	0.85	0.84	0.87	0.84	0.87
XML RoBERTa Near	**0.86**	**0.85**	**0.88**	**0.85**	**0.88**

Table 5.4: Classifiers metrics for Spanish language augmented datasets. Higher scores indicate better performance. Highest values per column in bold.

In the case of the Spanish language, as seen in Table 5.4, we also see that models trained on datasets augmented using distant languages performed worse or similar to the original models described in Section 5.1.1. However, we can see a substantial improvement in the metrics of the models trained on datasets augmented using near languages. The original Spanish training dataset was considerably smaller than the English dataset (see Section 4.2.1). This factor likely had an impact on the diversity of the Spanish dataset, which in turn was greatly improved by the synthetic variations produced in the backtranslation process. This finding can be useful to researchers, dealing with small datasets of underrepresented languages.

The results of the data augmentation study showed that using Transformers-based backtranslation to augmented text can produce very good results, especially if the original dataset is

relatively small.

The data augmentation study results show also an unexpected result: using distant languages for backtranslation produce greater variability, but that did not result in greater model accuracy. Actually it resulted in worst accuracy results.

Reviewing some of the examples of backtranslation from distant languages in Table 5.5 we can see evidence that some of the semantic content is indeed lost in translation.

Original English	Backtranslated from Chinese
I didn't feel humiliated	I'm not ashamed.
I can go from feeling so hopeless to so damned hopeful just from being around someone who cares and is awake	I can move from feeling so desperate to hope so desperate simply because of the people around me who care and wake up.
I'm grabbing a minute to post I feel greedy wrong	It took me a minute to post it and I felt greedy and wrong.
I am feeling grouchy	I don't feel well.
I've been feeling a little burdened lately wasn't sure why that was	I feel a little overstretched lately. I don't know why.

Table 5.5: Backtranslation examples from the English-Chinese-English pipeline

Whether the loss in the semantic content is due to deficient translation models or to some inherent difficulty in the translation between two distant languages is a question that is beyond the scope of this study. However, it is clear that, at least for now, the use of backtranslation for text data augmentation increases the performance of transformers-based models only with languages that at least share the same character set.

Finally, and based on the results of this data augmentation study, we can conclude that the transformers-based model to be used for emotion classification of the English tweets data is RoBERTa-base augmented with near languages backtranslation, and XLM (Multilingual) RoBERTa-base augmented with near languages backtranslation for Spanish tweets data. The performance gap between RoBERTa and its competitors is most likely explained by the differences in training data and hyper-parameters optimization RoBERTa has over BERT and DistilBERT. Being trained on a corpus ten times larger than the one used in BERT, helped RoBERTa learn more nuanced word relations tha its competitors. This allowed it to better detect nuances between classes when faced with a classification problem as the one addressed in this thesis.

5.3 Aspect Identification

In the previous section I went through the process of selecting and fine-tuning an appropriate and accurate methodology to detect emotions in tweets, measuring the accuracy of the multiple algorithms and finally setting down on using the RoBERTa deep language model in its English and Multilingual versions for emotion detection. Such a straightforward testing and selection process was possible due to the characteristics of the "emotions detection" task. In this study I only analyse four basic emotions (anger, fear, sadness and joy), and the fact that there is already a reasonable amount of tweets training datasets labelled with those particular emotions facilitated the use of supervised machine learning techniques for the detection of those four emotions in the tweets data of the social movements.

As stated in Section 4.4, in this thesis I'm using several unsupervised machine learning methods (LDAs, Bigram Networks) along with other knowledge discovery techniques to identify the four aspects (group-based motivator, collective actions, ingroup and outgroup) in the tweets corpus. In this section I show the results of the process of selecting the keywords for each aspect in each of the social movements datasets.

Below I present the evolution of the perplexity score (see 4.4) of LDA results using two to ten topics in each of the social movements tweets sampled datasets.

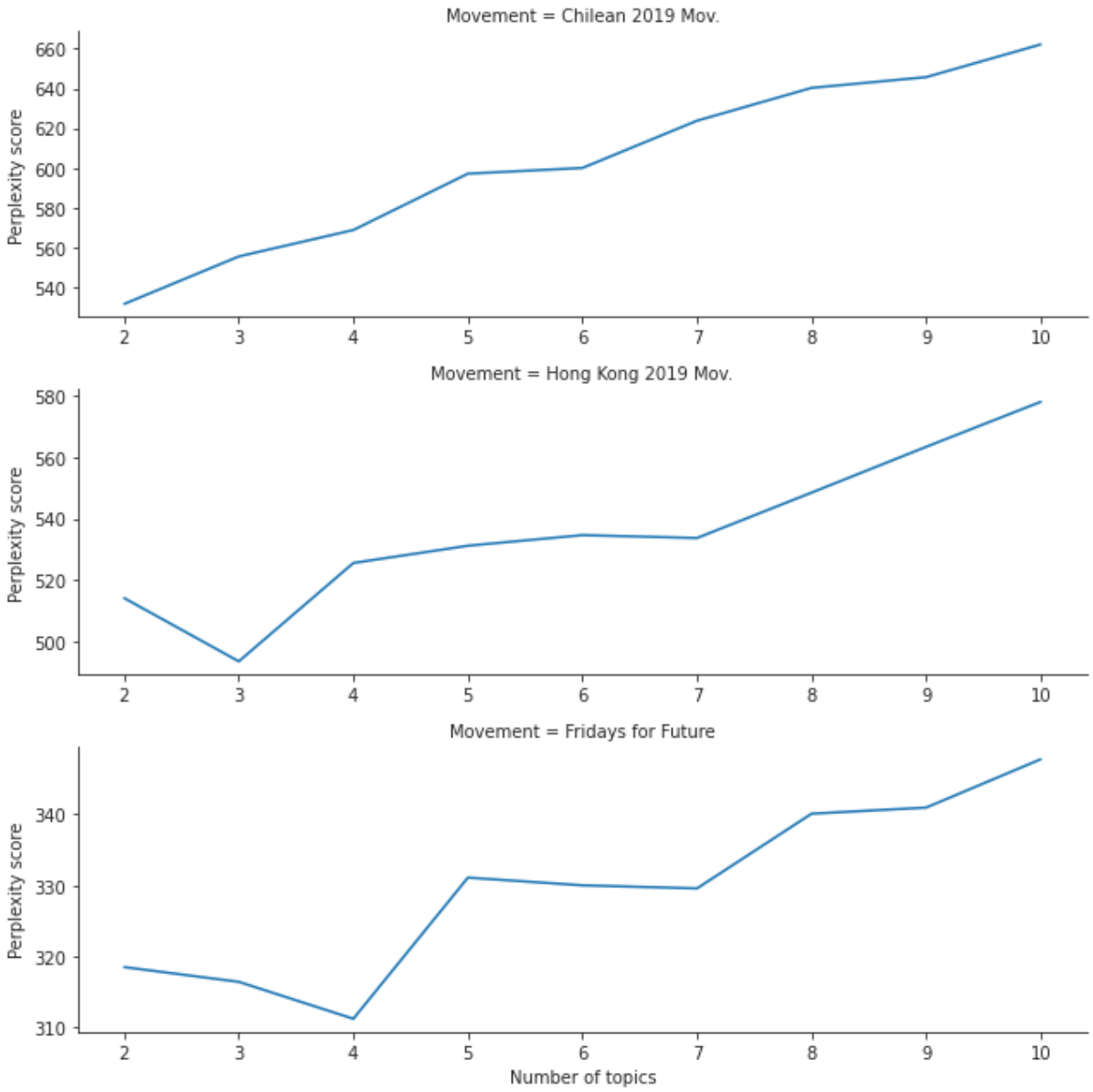

Figure 5.4: Top: Evolution of perplexity score of the Chilean 2019 Social Movement sample; Middle: Evolution of perplexity score of the Hong Kong 2019 Social Movement sample; Bottom: Evolution of perplexity score of the Fridays for Future Social Movement sample.

As was explained in Section 4.4 the perplexity score in LDA can be used as a way to assess how well the LDA model fits the real distribution of words in the corpus. A lower perplexity score means a better fit to the corpus, and hence a better model. Perplexity scores are specific to each corpus, so no comparison can be made across social movements, and the analysis is limited to determine which number of topics produce the best fit to the corpus of each social movement.

Figure 5.4 shows that for the Hong Kong 2019 movement and Fridays for Future, 3 and 4 topics, respectively, produce the best results. In the case of the Chilean 2019 movement, no number

of topics above 2 produce good fit. I cannot assume, however, that these 2, 3 or 4 topics are related to the four aspects I wish to identify as it is completely reasonable to think that the general conversation of the social movements was not structured around those four aspects. Here I am not using these topic solutions to extract the keywords for the four aspects, but as a way to inspect and visualise the corpus by reorganising it according to number of topics that best fits its internal structure, in the hope that this visualisation helps understand how the social movements defines its internal aspects.

5.3.1 Keyword Selection for the Chilean 2019 Social Movement

Here I will present the results of the word frequency, LDA and bigram network analysis and the final selection of keywords per aspect for this social movement.

Figure 5.5 shows the word frequency distribution in the Chilean 2019 Social Movement tweets corpus. As can be seen in the Figure, many of the most mentioned words are related to the government of Sebastián Piñera, mentioning him directly as "presidente" (president) or as "gobierno" (government). Other relevant mentions include the "constitucion" (constitution), "pais" (country), "pueblo" (people), "gente" (people synonym), "paronacional" (national strike), "protestas" (protests), "chilenos" (Chileans) and "carabdechile" (Carabineros de Chile, Chilean police force), among others. Even with this simple analysis, we can start to see the words that might fit into the categories of the four aspects of the Chilean Movement. As was described in Chapter 3, the Chilean 2019 movement had the authorities of Piñera's government as its main antagonist and it makes sense that he and his government are among the first ones mentioned. There is also several mentions of the collective actions performed by the movement. "Paronacional," "protestas," "cacerolazo" (a form of pan banging protest) are examples of collective actions which are commonly present in the corpus. Ingroup mentions at the top of the list include "gente," "pueblo," "chilenos," while mentions of the Group-based Motivator include "constitucion," "violencia," "ddhh" (Human Rights acronym) and "represion,"

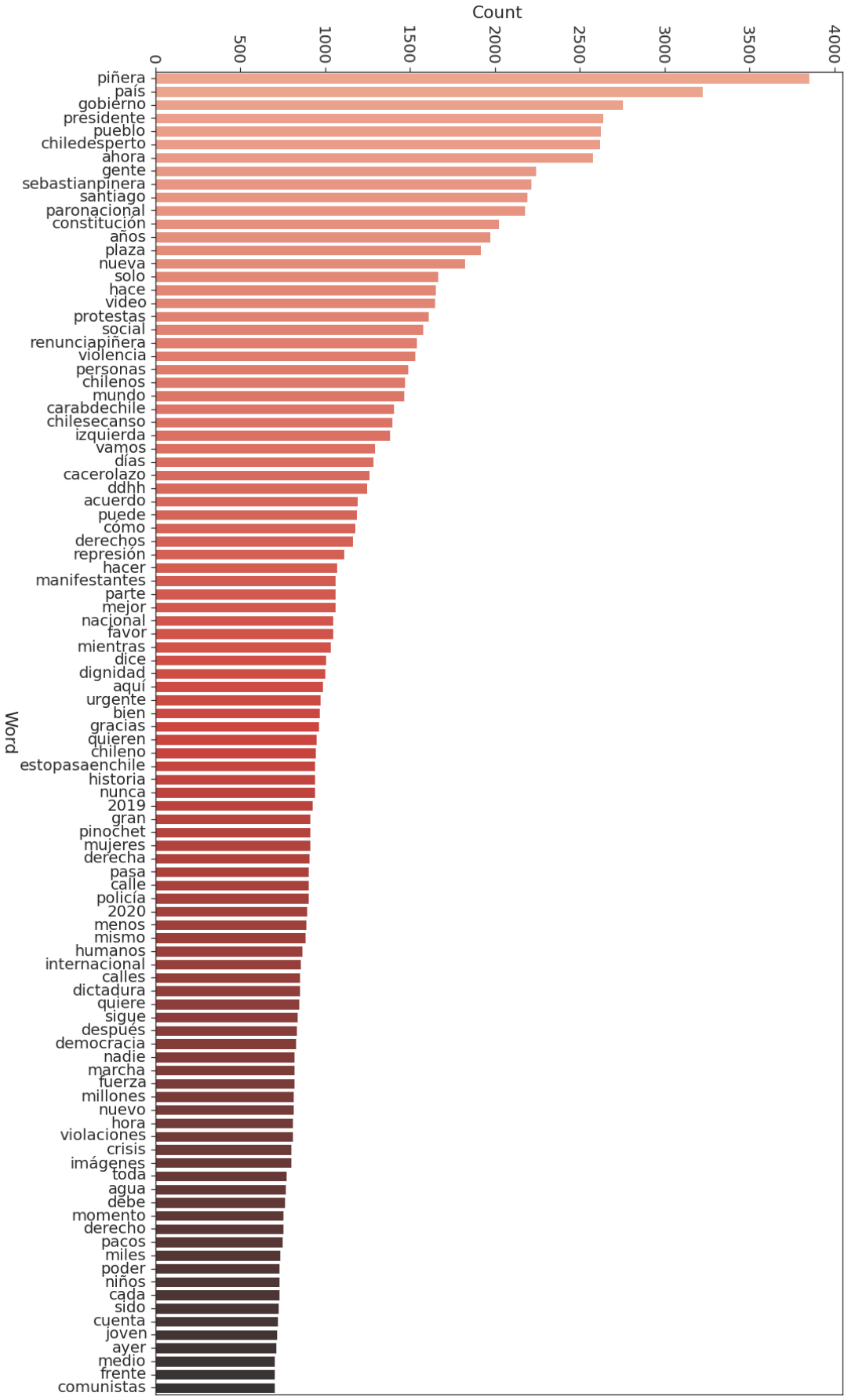

Figure 5.5: Frequency of the top 100 words in the Chilean 2019 Social Movement tweets sample.

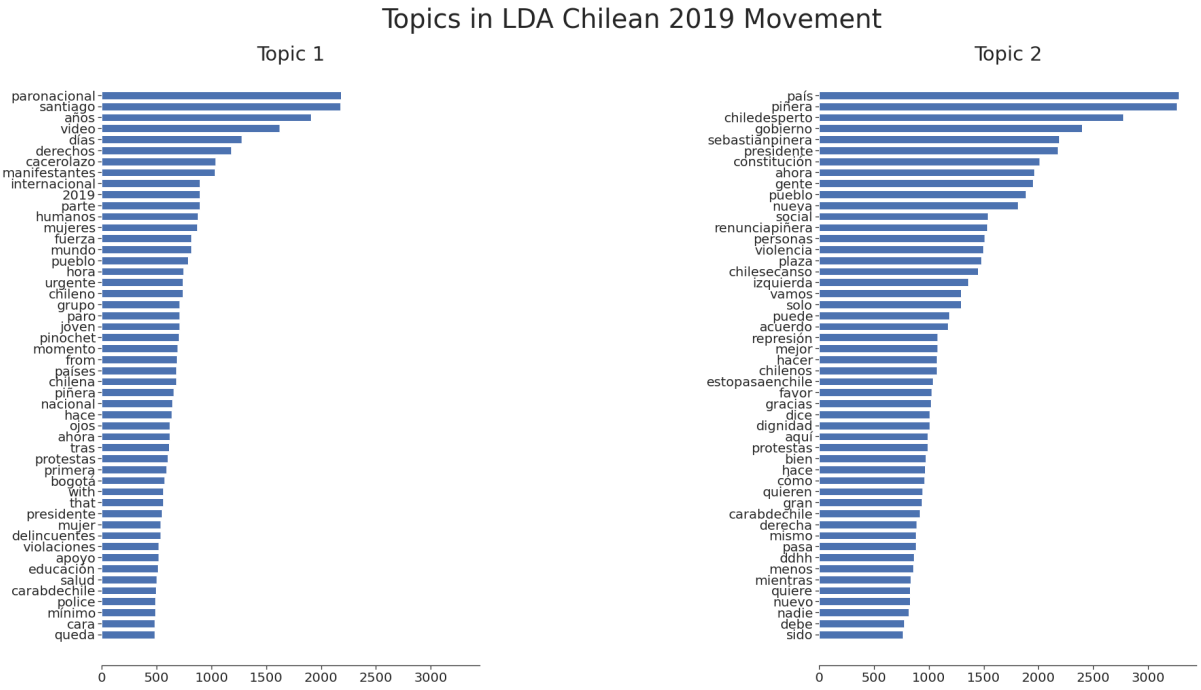

Figure 5.6: Latent Dirichlet Allocation topic modelling of the Chilean 2019 Social Movement tweets sample, showing the top 50 words per topic.

Figure 5.6 shows the results of the topic modelling of the Chilean 2019 Social Movement sampled corpus. The two topics extracted seem to show very similar words to the ones presented in Figure 5.5, however, the topic's composition is noteworthy. It seems that Topic 1 refers more prominently to Collective Actions and the Ingroup of the movement, with terms like "paronacional," "cacerolazo" and "manifestantes" (protesters) at the very top of the list, while Topic 2 seems to be focusing on the Outgroup and Group-based Motivator, with terms like "gobierno," "piñera," "sebastianpiñera," "presidente," "chiledesperto" (Chile woke-up), "nueva" (new), "constitucion" (constitution), "renunciapiñera" (resign piñera) in the top positions. Locations, such as "santiago" (Chile's capital) are also mentioned, but it is very likely these words were there to provide context to the tweet. For example, most of the massive protests occurred in the city of Santiago. Table 5.6 shows examples of tweets containing references to "Plaza Dignidad" (Dignidad Square).

Example tweets	Translation
Día 26 de protestas en Plaza de la dignidad (antes Plaza Italia)...	Day 26 of protest in Dignidad Square (former Italia Square)...
Estábamos en plaza dignidad ahora a las 23:25 de la noche...	We were in Dignidad Sq. now at 23:25 pm...
Plaza de la dignidad, sector Ramón Corvalán, pleno centro de Santiago...	Dignidad Sq., Ramon Corvalan sector, right in the centre of Santiago...
Plaza de la dignidad sigue con una multitud hermosa...	Beautiful crowd still in Dignidad Sq...

Table 5.6: Examples of Chilean 2019 tweets mentioning "dignidad."

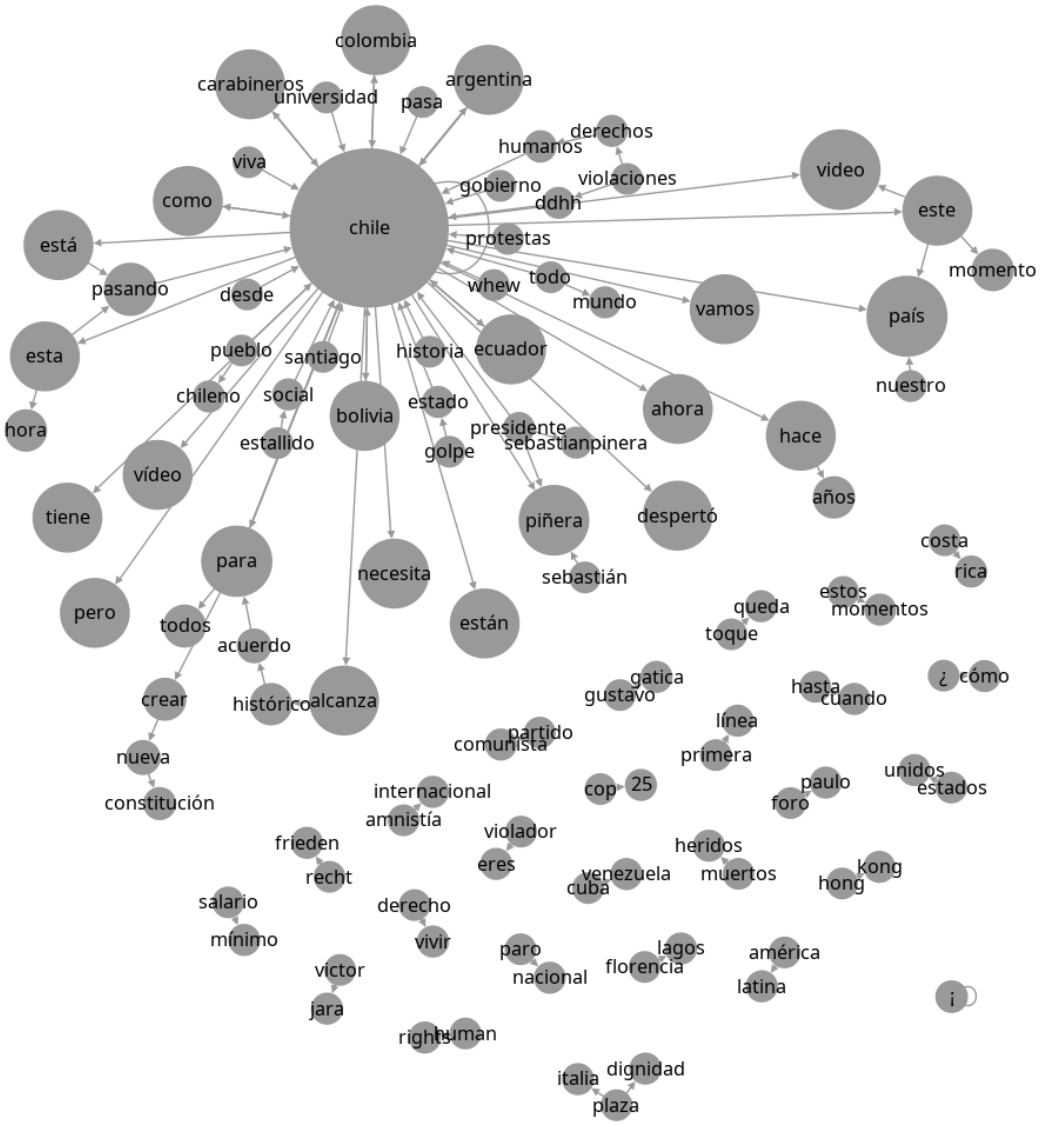

Figure 5.7: Network of the 100 most common bigrams found on the Chilean 2019 Social Movement tweets sample. Image produced using *Gephi* graph visualisation tool.

Figure 5.7 shows the bigram network of the most common 100 bigrams extracted from the Chilean 2019 Social Movement sampled corpus. The size of the nodes of the network is determined by its eigenvector centrality value[1]. In Figure 5.7 we can see that "chile," unsurprisingly, is the term with the greatest importance, having connections to many other terms. Noteworthy configurations of bigrams are "violacion," "derechos," "humanos," "chile" (violation, human, rights, Chile) located to the right of the "chile" node. Tweets containing this sentence were a way to denounce the excessive use of force by the police. The connection of the terms "chile," "alcanza," "historico," "acuerdo," "para," "crear," "nueva," "consitucion" (Chile reaches agreement to create new Constitution) was also a very common sentence mentioned during the protest period, highlighting one of the main concession achieved by the movement, reflecting that a rewriting of the constitution was an important motivator for the protesters. Table 5.7 shows examples of tweets containing references to the rewriting of the constitution and the Human Rights violations occurred in Chile during the protests, which in turn are examples of the movement's group-based motivators.

Example tweets	Translation
Chile camino a una nueva constitución y un cambio de modelo de país...	Chile on its way to a new constitution and a change in its country model...
El pueblo de Chile se ganó el derecho a escribir la constitución que lo rige...	The people of Chile earned the right to write a new constitution...
El pueblo de Chile destruyendo la constitución del 80. Uno de los muchos horrocruxes de Pinochet...	The people of Chile destroying the 1980 constitution, one of the 'horcruxes' of Pinochet...
Violando los Derechos humanos, la policía de Chile sigue abusando...	Violating Human Rights, the Chilean police continue its abuse...
Denuncian graves violaciones a los derechos humanos en Chile	Serious Human Rights violations denounced in Chile...
En Chile se violan los derechos humanos!!! Gustavo Gatica joven de 21 años perdió sus dos ojos en las manifestacion...	Violation of Human Rights in Chile!!! A young 21 yo man, Gustavo Gatica, lost his eyes in the protests...

Table 5.7: Examples of Chilean 2019 tweets mentioning "constitucion" and "derechos humanos."

[1]The eigenvector centrality is a measure of the influence of the node over the entire network. A higher value of this measure indicates that the node is connected to many nodes that also have high eigenvector centrality.

Group-based Motivator	Collective Actions	Ingroup	Outgroup
chiledesperto (chilewokeup)	protestas (protests)	pueblo (people)	Piñera
chilesecanso (chilehadenough)	protesta (protest)	gente (people)	piñera
dignidad (dignity)	paro (strike)	izquierda (left)	sebastian
represión (repression)	paronacional (nationalstrike)	pueblo chileno (chilean people)	presidente (president)
ddhh (human rights)	marcha (march)	estudiante (student)	carabineros (police)
derechos humanos (human rights)	marchas (marches)	estudiantes (students)	carabinero (police)
constitucion (constitution)	cacerolazo (pan banging)	manifestantes (protesters)	gobierno (government)
nueva constitucion (new constitution)	plaza dignidad (dignity square)		sebastianpinera
constitución (constitution)			sebastianpinera
derechos (rights)			carabdechile (police)
pinochet			piñerarenuncia (piñeraresign)
			renunciapiñera (resignpiñera)
violaciones (violations)			

Table 5.8: Chilean 2019 Social Movement selected aspect keywords. Translations in parenthesis.

Finally, Table 5.8 presents the final selection of keywords representing the fours aspect of the Chilean 2019 Social Movement, based on top most frequent words, combined with the top words in the topic modelling and the central words in the bigram networks. All the results help me triangulate the most important terms. The keywords of the Ingroup and Outgroup aspects are possibly the most self-evident and certainly the most common in the analyses presented. The government of Piñera was undoubtedly the main antagonist of the movement, followed closely by the Chilean police force: "carabineros," which although was hidden away from the analysis given its extremely high frequency in the corpus, was still included in the final Outgroup keywords. The people ("gente," "pueblo"), the protesters ("manifestantes") the students ("estudiantes") and the left ("izquierda") were widely regarded as the core participants of the movement. Mentions of them in a tweet should suggest mentions of the Ingroup. The Collective Actions keywords are all mentions of specific forms of protest, with the exception of "plaza dignidad." This term was included only because during the protest of 2019 the mere mention of

this specific place, located in the city of Santiago de Chile, was synonymous with protests and other forms of collective action. We can also note that mentions of violent collective actions are not present, the only terms that could be related to that is "violencia" (violence). However, that term was ambiguous in the conversation as it was used to refer to the police violence more often than to the protesters violence. Additionally, there seems to be a tacit agreement in the conversation, in which protesta (protest) could mean violent protest or peaceful protest equally often. Finally, the keywords selected to represent the Group-based Motivator are a mix of battle cries ("chiledesperto," "chilesecanso") (Chile woke up, Chile got tired) which represented the general feeling of outrage of the protesters, mentions of human rights violations ("violaciones," "ddhh," "derechos humanos"), repression ("represion") and hopes of writing a new constitution, leaving behind the old one made during the dictatorship of Pinochet. Table 5.9 shows tweets examples of the four aspects identified in the dataset.

Example tweets	Translation
Group-based Motivator	
Chile: un reclamo por las faltas de derechos esenciales... Chile avanza. La Constitución nacida de la dictadura de Pinochet llega a su fin... Es un momento histórico. La ciudadanía decidirá cómo se conformará la instancia que redactará la nueva Constitución...	Chile: A cry becasue of the lack of essential rights... Chile moves forward. The dictatorship born constitution of Pinochet comes to and end... Historial moment. The people will decide how the organisation in charge of writing the new constitution will be composed...
Collective Actions	
El cacerolazo está aplastando a los vándalos, sonando más fuerte que ellos, y reivindicando nuestro derecho a protestar. Las justas protestas sociales en contra del modelo neoliberal de explotación capitalista se extienden por América Latina. Se armó: Convocan a la marcha K-Pop más grande de Chile para este viernes.	The "cacerolazo"[2] is crushing the vandals, sounding louder than them and reivindicating our right to protest. The just social protests against the exploitative capitalist neo-liberal model spread throughout Latin America. It's on: Call for the biggest Kpop march of Chile this Friday.
Ingroup	
Un grupo de manifestantes derribó la estatua del conquistador español, Pedro de Valdivia... Vean, el pueblo chileno masivamente en las calles contra las declaración de guerra de Piñera...	A group of protesters toppled the Spanish conqueror statue of Pedro de Valdivia... Look, the masses of people of Chile in the streets against the declaration of war of Piñera...
Outgroup	
No estoy de acuerdo con la convocatoria que ha hecho el Presidente al COSENA. Chile NO está en Guerra. Gobierno dice q funar la PSU atenta contra D° a la Educación, les informamos q en Chile no existe el derecho a la educación Aprobación de Piñera cae a un 9,1%. ¡Se consolida como el Presidente peor evaluado en la Historia de Chile!	I'm against the president summoning the COSENA[3]. Chile is NOT at war. The government says that boycotting the PSU[4] goes against the right to education, we inform them that in Chile there is no right to education. Approval rate of Piñera fall to a 9,1%. It consolidates him as the worst rated president of the history of Chile!

Table 5.9: Examples of Chilean 2019 tweets for each socio-psychological aspect.

5.3.2 Keyword Selection for the Hong Kong 2019 Social Movement

As it was the case with the previous section, here I will present the analysis leading to the selection of the keywords that best represent the four aspects (Group-based Motivator, Collective Actions, Ingroup and Outgroup) in the tweets related to the Hong Kong 2019 social movement.

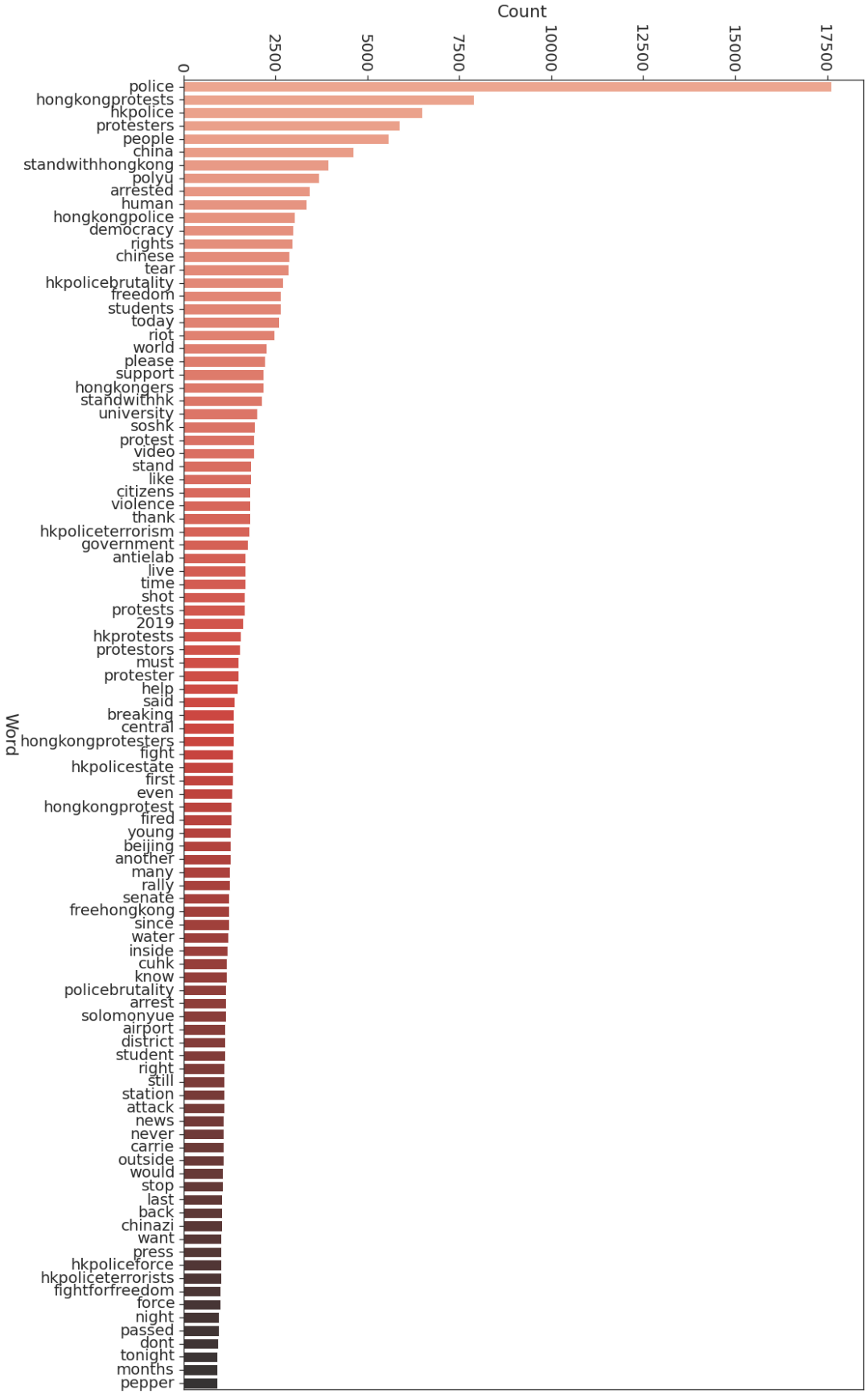

Figure 5.8: Frequency of the top 100 words in the Hong Kong 2019 Social Movement tweets sample.

Figure 5.8 shows the word frequency distribution in the Hong Kong 2019 Social Movement sampled corpus. We can see a high count of terms referring to the Hong Kong police forces ("police," "hkpolice"), one of the main antagonist of the social movement. Among the top words we can also observe mentions of the "protesters" and to the "hkprotests," keywords related to the Ingroup and Collective Actions of the movement, respectively. Below these top keywords, we see the first mention of "china," one of the parties seen as the antagonist/Outgroup of the movement. After that we see mentions of "polyu" (The Hong Kong Polytechnic University), one the Universities sieged by the Hong Kong police during the 2019 protests, and mentions of some of the demands/Group-based Motivators of the movement: "democracy," "human rights," "hkpolicebrutality," "freedom."

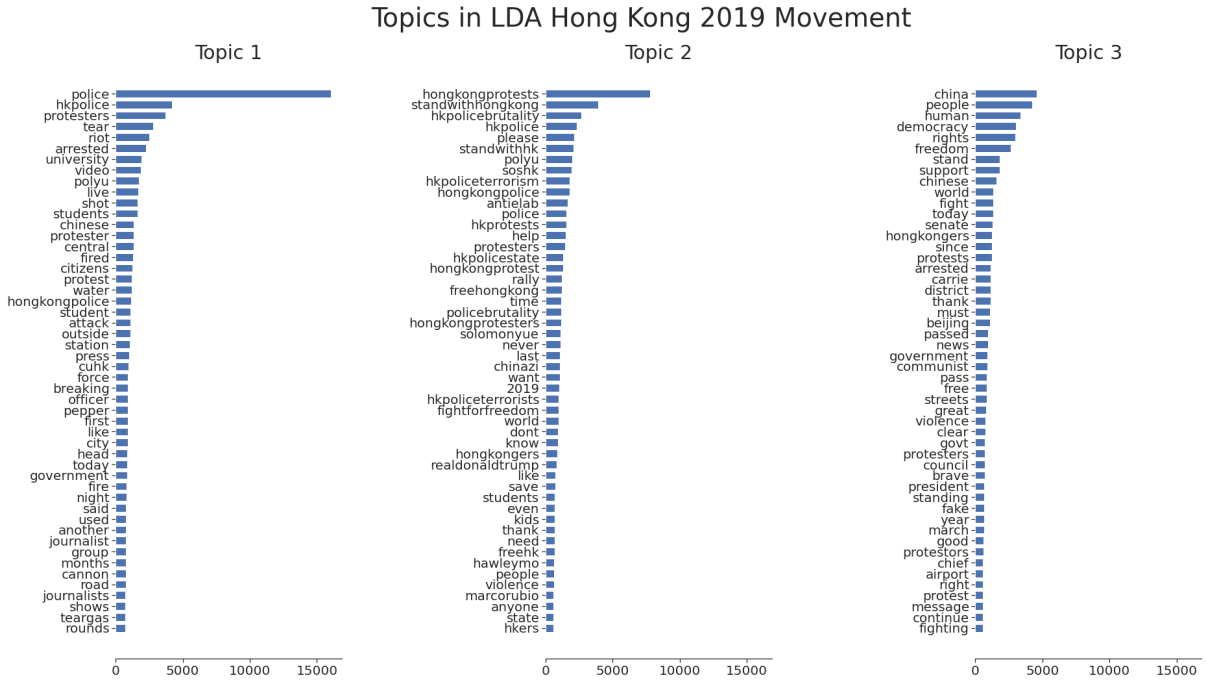

Figure 5.9: Latent Dirichlet Allocation topic modelling of the Hong Kong 2019 Social Movement tweets sample, showing the top 50 words per topic.

Figure 5.9 shows the results of the topic modelling of the Hong Kong 2019 Social Movement data sample. This time, and following the results of the analysis on the optimal number of topics shown in Figure 5.4, three topics were extracted. In general terms Topic 1 seems to be focusing on the students and protesters of Hong Kong and the reaction of the Hong Kong police. Topic 2 seems to have focused only on the hashtags related to the movement. Finally, Topic 3 seems to be focusing on the international actors involved, such as China and the US Senate, and the general motivations of the movement, such as their demands for democracy and freedom. The

topic modelling analysis has some concordances with the raw frequency count, with mentions of keywords representing the four aspects in the top counts of topics 1 and 3.

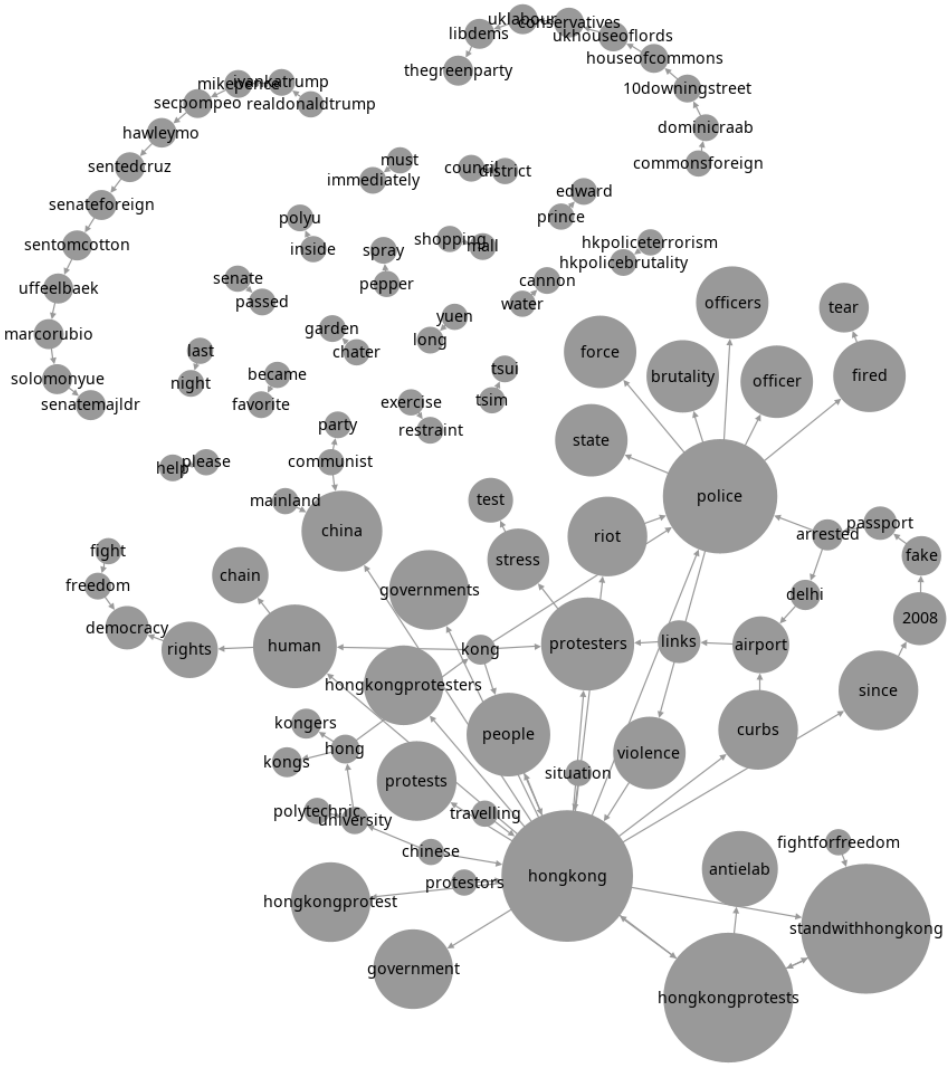

Figure 5.10: Network of the 100 most common bigrams found on the Hong Kong 2019 Social Movement tweets sample. Image produced using *Gephi* graph visualisation tool.

Figure 5.10 shows the bigram network of the top 100 bigrams extracted from the Hong Kong 2019 Social Movement sampled corpus. Again, the size of the node is determined by its eigenvector centrality value, which can be interpreted as the importance or "prestige" of the node inside the network. Within the network there are several interesting chains of bigrams related to the aspects of the social movement. The connection of the term "hongkong" to the terms "police" which is immediately related to the terms "brutality," "tear (gas)," "force" and "violence" speaks about the police brutality seen during the protest and which was one the main motivations of

the movement. The term “hongkong” also relates to the term “china” which then is connected to “mainland” and “communist party,” revealing that these two entities were mentioned in the online conversation and that they were likely regarded as antagonists of the movement. Finally, the last noteworthy chain of bigrams relate “hongkong” with “human rights,” “democracy” and “freedom,” all values which served as main motivators for the participants of the movement. Table 5.10 shows tweets examples containing mentions of ‘police,” ‘china,” the ‘ccp’ and ‘human rights’.

Example tweets
The white vans of the police ran INTO the protesting crowd and caused a stampeding accident which led to numerous injuries.
HK police blocked main roads in the busy commercial area Central and fired rounds of tear gas.
CCP Army is in the HK Police Force. Cantonese is spoken in HK, while this man dressed in police uniform speakes fluent Mandarin. #HongKongProtests #LiberateHongKong #chinazi.
You’re right! Chinese are always make #doublestandard. #hongkong #protesters are not riots compared with those in #china. Or the real rioters are #hk #police, they all have the weapons to kill or attack anyone. #antiELAB #freehongkong
Here we are blocking the tomb of Mao Zedong standing for human rights in Tiananmen Square during the Beijing Olympics.
I’m proud to stand in solidarity with the people of #HongKong as they stand up for their basic human rights.

Table 5.10: Examples of Hong Kong 2019 tweets mentioning ‘police,” ‘china’, ‘ccp’ and ‘human rights’.

Group-based Motivator	Collective Actions	Ingroup	Outgroup
freedom	hongkongprotests	young	China
democracy	hongkongprotest	students	china
human right	hkprotests	citizens	ccp
freehongkong	strike	people	chinese comunist party
freedomhk	protest	protesters	mainland
		protester	police
		hkprotesters	hkpolice
		hongkongprotesters	hongkongpolice
		hongkonger	goverment
		hongkongers	
		hongkongner	
		hongkongners	

Table 5.11: Hong Kong 2019 Social Movement selected aspect keywords.

Finally, Table 5.11 presents the final selection of keywords representing the four aspects of

the Hong Kong 2019 Social Movement. This selection was made taking the top words in the frequency analysis, the top words in the topic modelling and the central terms in the bigram networks analysis performed above. The Group-based Motivator keywords correspond to the main concerns and demands of the movement (freedom, democracy, human rights). In the Collective Actions keywords list I included mentions of strikes and protests (and several related hashtags). Again, as it was the case with the Chilean 2019 Social Movement, the corpus did not include mentions of violent collective actions. This is mainly because most of the Twitter campaigns pro Hong Kong were in English, but most of the criticism was done in Chinese, mainly driven by the Chinese Government (DW 2019). Although this is a structural bias of the data collection process, for the case of this study is rather beneficial, since it allows to capture the opinion and emotions of mainly the supporters of the movement. The Ingroup keywords correspond to mentions to the citizens of Hong Kong, the protesters and the "young" "students" who participated in the defence of the Hong Kong Universities. I additionally included several hashtags referring to the protesters and the Hongkongers. This decision was also influenced by the fact that the pro Hong Kong campaign was mainly done in English, and mentions to citizens of Hong Kong were considered mentions to the Ingroup. Finally, in the Outgroup I included mentions of the Hong Kong police, the Hong Kong "government," mentions to Mainland China, and to the Chinese communist party, all of which were found to be present in the topic modelling and network bigram analyses. Table 5.12 shows examples of tweets related to the four aspects of the Hong Kong 2019 movement.

Example tweets
Group-based Motivator
Thanks to US House and all Americans for supporting Human Rights and Democracy... whole world must wake up now ! Proud of HongKong peaceful protesters and bravely students who fight for human rights and freedom... respect you all - world heroes Today marks an important step forward in the fight for freedom and human rights around the globe. Here I am in Chater Garden in central Hong Kong standing and praying for democracy, freedom and human rights.
Collective Actions
In a Peaceful Protest Event in TST[5] today. An Riot Mobs- Polices come out and spray a pepper spray. NOW: #HongKongProtests continue with marches happening across the city in support of the pro-democracy movement. We urge everyone in Hong Kong to join our protest on coming Sunday, to support human rights day 2019.
Ingroup
In plea to former colonial power, #HongKong protesters sing 'God save the queen'. Resist now or we can never resist. Please support CUHK students and Hong Kongers. Protesters rush away from a section of Nathan Road after the water cannon truck fires blue liquid.
Outgroup
HK is police state under rule of CCP. I cannot believe they attack the muslim temple. Ok this is SUPER WORRYING. Where are #HongKong #Police taking the arrested #hongkongers to, via a train? To #China? Police threatens trying to push back journalist using their batons and weapons.

Table 5.12: Examples of Hong Kong 2019 tweets for each socio-psychological aspect.

5.3.3 Keyword Selection for Fridays for Future

Finally, here I'm presenting the analysis leading to the selection of the keywords that represent the four aspects (Group-based Motivator, Collective Actions, Ingroup and Outgroup) in the tweets related to the Fridays for Future movement.

Figure 5.11 shows the word frequency distribution in the Fridays for Future movement sampled corpus. As is expected for a movement concerned with climate change, the top words refer immediately to climate, one of the main Group-based motivators of Fridays for Future (e.g., "climate," "4climate," "climatechange"), followed by mentions of collective actions ("school-strike," "strike") and mentions of the founder of the movement Greta Thunberg. Bellow the top 10 we start seeing mentions of the Ingroup ("young," "student," "youth," "kids"). Finally,

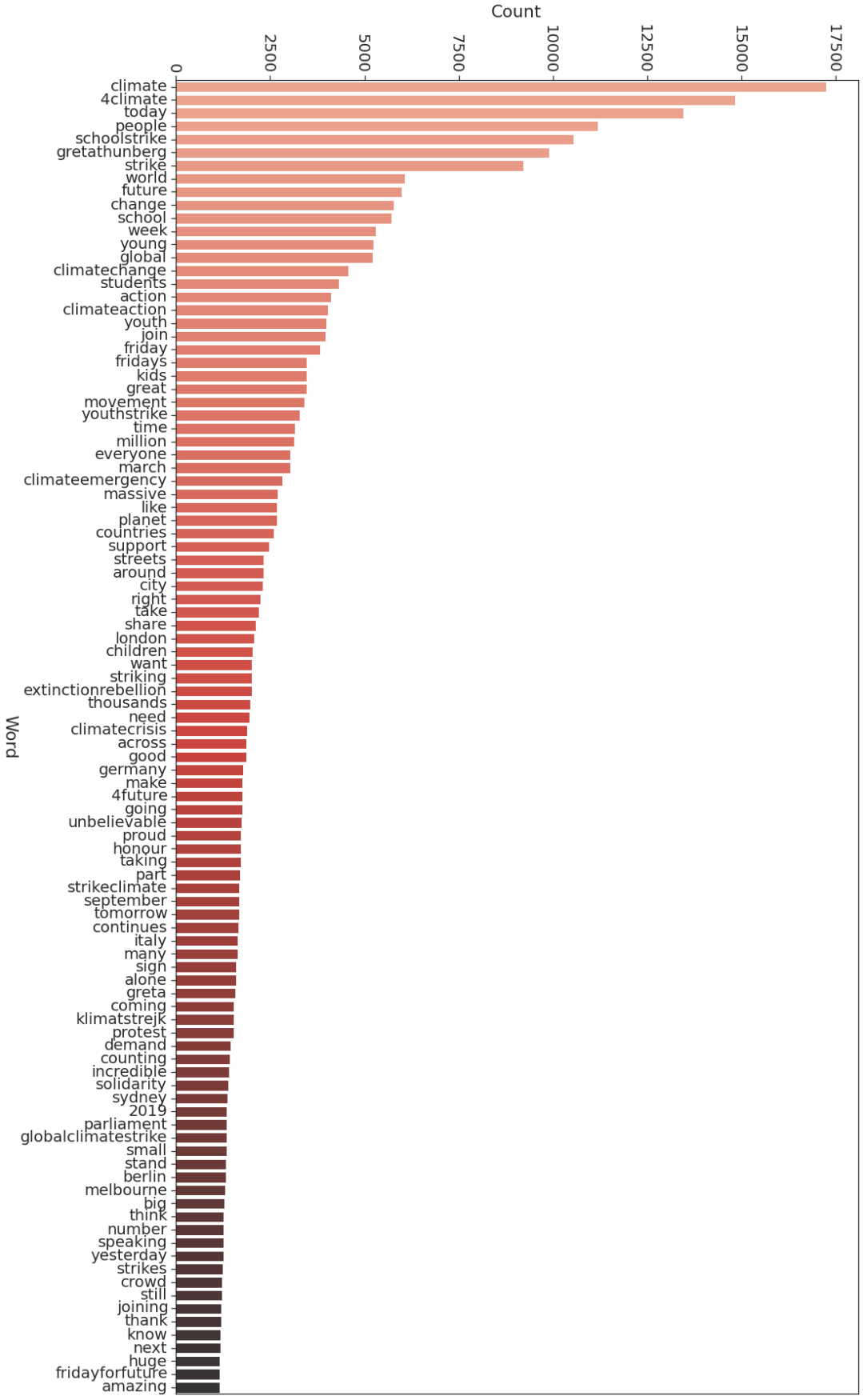

Figure 5.11: Frequency of the top 100 words in the Fridays for Future tweets sample.

around 80th position we see the first mention of one of the declared Outgroup/antagonist of the movement: the "parliament."

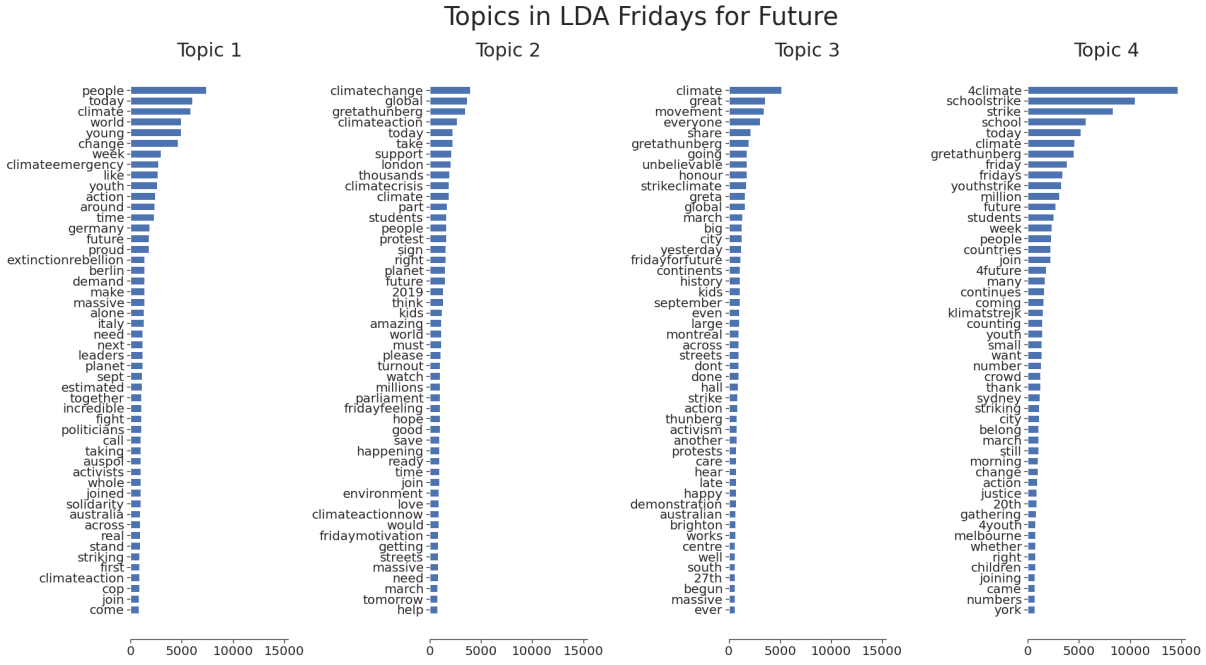

Figure 5.12: Latent Dirichlet Allocation topic modelling of the Fridays for Future tweets sample, showing the top 50 words per topic.

Figure 5.12 shows the results of the topic modelling extraction over the Fridays for Future sample. Following the results of the analysis on optimal number of topics shown in Figure 5.4 I fitted a model assuming the presence of four topics in the corpus. The top words are mainly related to the climate crises, the school strikes and Greta Thunberg. Below the top five words, the topics start to show differences between them. Topic 4 seems distinctly more concerned with the Collective Actions and the Ingroup of the movement featuring words like "schoolstrike," "strike," "youthstrike" and "students" relatively high in its list. Topic 1 and 2, on the other hand, seem to be focusing on the Group-based Motivator of the movement, with mention to the "climate," "climateemergency," "climatechange," and "climatecrisis." Finally, Topic 3 seems to be focusing on the figure of Greta Thunberg and the "School Strike for Climate" actions. Mentions to the Outgroup occur in Topics 1 and 2, relatively low on the list ("leaders" in position 26th of Topic 1, "politicians" in position 33th of Topic 1, "parliament" in position 30th of Topic 2).

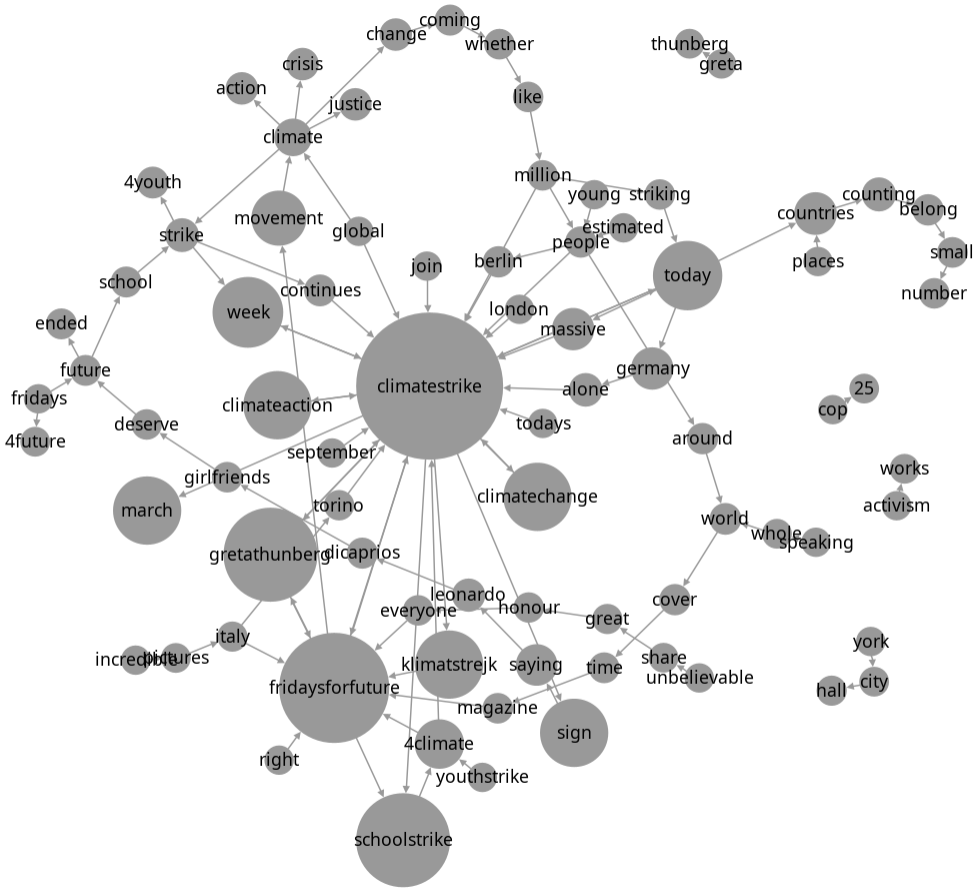

Figure 5.13: Network of the 100 most common bigrams found in the Fridays for Future tweets sample. Image produced using *Gephi* graph visualisation tool.

Figure 5.13 shows the bigram network of the top 100 bigrams extracted from the Fridays for Future sampled corpus. We can see that the main node is "climatestrike" (a mention of a Collective Action) at the centre of the network. Other important nodes are "gretathunberg" (an Ingroup member), "fridaysforfuture" (a mention of the Ingroup), "schoolstrike" (a mention of a Collective Action), "climatechange" (a Group-based motivator) and "climateaction." Connections of the word "climate" reveals also that concepts like "climate crisis" and "climate justice" are concerns for the movement. The connection of the token "fridaysforfuture" to "movement" and subsequently to "climate" and "justice," reinforce the idea that the movement is committed to the idea of pushing the agenda of "climate justice." Table 5.13 show examples of Fridays for future tweets containing some of these terms.

Example tweets
We are preparing to strike for the climate here in Karlskrona! My placard is locked and loaded with this lovely section of a poem by lemnsissay (Lemn Sissay).
Climate change is happening, humans are causing it, children are are fixing it! #FridaysForFuture protest in Kilkenny...
Today's #ClimateStrike is in proud tradition of young people striking for justice. 1972 school strike against corporal punishment, 1980s black student strikes against racist violence, 2003 school strikes against War in Iraq.
Sustainable development is the pathway to the future we want for all. It offers a framework to generate economic growth, achieve social justice, exercise environmental stewardship and strengthen governance.
By investing hundreds of millions of dollars in fossil fuels and over $91 million in Israeli banks and Elbit Systems, AXA is financing the climate crisis and Israeli apartheid.
Injustice and greed created the climate crisis. The only way we can fight back is by organizing for a just and sustainable world. And it's going to take all of us.

Table 5.13: Examples of Fridays for Future tweets mentioning 'climate', 'justice', 'crisis'.

Group-based Motivator	Collective Actions	Ingroup	Outgroup
climatechange	climatestrike	fridaysforfuture	leaders
climate change	climate strike	fridays for future	politicians
climatecrisis	schoolstrike	fridays4future	parliament
climate crisis	school strike	kids	
climateemergency	schoolstrikeforclimate	youth	
climate emergency	school strike for climate	students	
climate	schoolstrike4climate	movement	
globalwarming	schoolstrike 4climate	children	
global warming	civil disobedience		
pollution	strike		
carbon emissions	protest		
carbon footprint			
carbon			

Table 5.14: Fridays for Future selected aspect keywords.

Finally, Table 5.14 presents the final selection of keywords representing the fours aspect of Fridays for Future. The keywords representing the Group-based motivator are mainly composed of variations of "climate change," "climate emergency/crisis," "global warming" and "carbon emission," all of them predominantly present in the tweets corpus and/or in the officially declared motivation and objectives of the movement. The keywords representing the Collective Actions of the movement are variations of the terms "climate strike" and "school strike." The Ingroup keywords are composed of variations of the term "fridays for future" and mentions to the "students," "youth," "kids" and "children" who are the ones comprising the bulk of the

participants of the movement. Finally, the Outgroup keywords are mainly composed of mention to authorities.

Table 5.15 shows examples of tweets related to the four aspects of the Fridays for Future movement.

Example tweets
Group-based Motivator
British Teenagers Want Radical Action On Climate Change And Say They're Not Taking No For An Answer! March for a better tomorrow with millions of people from 156 countries who will take to the streets in support of urgent action on climate change. Climate change is an existential threat—and we are already facing the effects.
Collective Actions
Get up, get out, get involved. Join the #ClimateStrike and help protect your world. So many people came to strike in israel today!! Pretty amazing! #FridaysForFuture #ClimateStrike. #ClimateStrike Belfast city centre. So much positivity, peace and passion in the city I love. Couldn't be prouder...
Ingroup
The amazing students at a high school just north of Seattle have a #FridaysforFuture #ClimateStrike every Friday. School children striking for meaningful #ClimateAction in #Helsinki today. The kids of EU are doing it today. They lead the world. #klimastrejke
Outgroup
It's great to see another huge #ClimateStrike heading towards the Scottish Parliament. Scotland needs a Green New Deal to transform our economy, cut emissions and create jobs! BREAKING - Four young people being arrested on Westminster Bridge #YouthStrike4Climate OccupyLondon. The police should be arresting the politicians for their criminal inaction on the climate and ecological crisis not the youth striking for their future.

Table 5.15: Examples of Fridays for Future tweets for each socio-psychological aspect.

The keywords selected in this section will now be used to filter the tweets of each of the social movements in order to inspect the emotions expressed by them in relation to the four aspects: Group-based motivator, Collective Actions, Ingroup and Outgroup.

5.4 Emotions in Social Movements

Having identified the most suitable methodological approach for the detection of emotions and the filtering of tweets based on the keywords related to the four social movements aspects, in

this section I will present the evolution of the emotions detected in the Twitter data collected for each social movement, using the previously described methodologies.

To recapitulate what was stated in Chapter 1, the objective of the classification is to detect four emotions (joy, sadness, anger and fear) across the tweets of the several social movements presented in Chapter 3. Moreover, I will test to what extent the four detected emotions mentioned above can predict the type of protest events observed for these three social movements. Each of these social movements has its own specific strategy in terms of collective actions. Fridays For Future is distinctively peaceful and non-violent in all of its activities, whereas the Chilean 2019 and Hong Kong 2019 social movements do have a strong component of violent civil disobedience. Given this, I expect different emotions to play a role in each of the social movements predicting violent or non-violent collective actions.

But before delving into these results, it is necessary to take a more detailed summary look at the data used in this study.

5.4.1 Descriptive Statistics of Twitter Data

Table 5.16 shows the number of tweets for each social movement containing terms related to each of the four aspects mentioned (group-based motivators (GBM), collective Actions, ingroup and outgroup) plus the number of "Ambiguous" (two or more aspects have the same count, see Section 4.1.1 for details) and "None" (zero mention to any of the aspects, see Section 4.1.1 for details) tweets.

Social Movement	GBM	Col. Actions	Ingroup	Outgroup	Ambiguous	None
Chilean 2019	1654904	1100906	1242672	2933644	1210512	14357973
Hong Kong 2019	747845	1007793	1975955	3440219	2240685	5264731
Fridays For Future	102522	802604	243702	1566	436204	34212

Table 5.16: Number of tweets per aspect by social movement.

In Table 5.16 we can see that the filtering of "Ambiguous" and "None" tweets reduced considerably the number of tweets per social movement. We can also observe that different aspects have different levels of significance among the five social movements studied. Table 5.17 shows the percentage of tweets per aspect by social movement.

Social Movement	GBM	Col. Actions	Ingroup	Outgroup	Ambiguous	None
Chilean 2019	7.35%	4.89%	5.52%	13.04%	5.38%	63.81%
Hong Kong 2019	5.10%	6.87%	13.46%	23.44%	15.27%	35.87%
Fridays For Future	6.32%	49.52%	15.03%	0.09%	26.91%	2.11%

Table 5.17: Percentage of tweets per aspect, by social movement. Note: The % was calculated using the number of tweets after filtering by ambiguous emotion classification

An important thing to consider at this point is the significant number of tweets classified as "None" or "Ambiguous" in the Chilean and Hong Kong movements. In particular, 63% of the tweets collected from the Chilean 2019 movements had no mentions to any of the aspects (Group-based Motivator, Collective Actions, Ingroup or Outgroup). A smaller, but still considerable 36% of tweets in the Hong Kong movement also did not contain any mentions of these aspects. This, although maybe disheartening, was not unexpected. The four aspects of the social movements were not intended to be an exhaustive classification of all the topics in the conversation around a social movement. They only reflect the aspects I consider relevant for their participation in collective actions, and hence non-relevant tweets were expected to creep in into the data collection. Additionally, an undetermined amount of unrelated tweets were expected to be collected, given that the keywords and hashtags, although specific for each movement, could be used by any user to jump into the trends, polluting the conversation with non-relevant tweets. Finally, another important cause of tweets falling into the "None" or "Ambivalent" classes is the fact that many tweets share other types of resources along with text (e.g., videos, images, or websites). Most of those tweets contain the relevant hashtags of the social movements and very short pieces of valid text (e.g., "This video bellow is so terrible," "the picture bellow makes me so angry," etc.), but the core of the content is presented in another format that is not analysed in this thesis.

On the other hand, very few tweets of the Fridays for Future movement did not contain any mentions of these aspects. Table 5.18 shows examples of tweets classified as "None" or "Ambiguous."

Examples of "Ambiguous" tweets
Hong Kong Hong Hum District Chief Inspector lost his temper and arrest innocent citizen. The men was actually wanted to rescue the students who fell from 3/F. So he tried to stop #HKPolice.
I'm a climate fighter". Listen to ten-year-old Parker recite a spoken word poem at the Brisbane #ClimateStrike
Examples of "None" tweets
Buzzing to be adding shows in Mexico and Chile to the tour! #chiledesperto
Dear #HK Do you think the international community should declare its support for universal suffrage in Hong Kong

Table 5.18: Examples of tweets classified as "Ambiguous" or "none."

5.4.2 Emotions in Social Movements

In the following section I will present the results of the emotion classification for each social movement. Given that tweets are time-tagged, I will present the emotion classification results in time-series fashion to capture the dynamic nature and evolution of emotions and their relation to real-life events. In order to provide a cleaner visualisation of the results, the data were aggregated by day, meaning that changes in emotions within a calendar day are smoothed out. This transformation allows to couple the emotion data with events data, which has a day-based report periodicity.

The presentation of the results will follow the order set in Chapter 3, with the Chilean Movement first, followed by the Hong Kong 2019 Movement, and finally Fridays For Future.

This presentation of results will include first a description of the general evolution of emotions of each social movement in relation to the reported events, to later dive into the evolution of emotions in relation to the events but separated by each aspect of the social movements described in Chapter 3, i.e., Group-based Motivation, Collective Actions, Ingroup, Outgroup.

5.4.3 Emotional patterns in the Chilean 2019 Social Movement

Overview

Figure 5.14 shows the evolution of the four emotions (sadness, anger, fear and joy) detected in the tweets data after filtering, described in Table 4.2. At the bottom half of the Figure, we see the reported events based on ACLED data. Finally, the transparent vertical bars of various colours represent moments of special interest in the data.

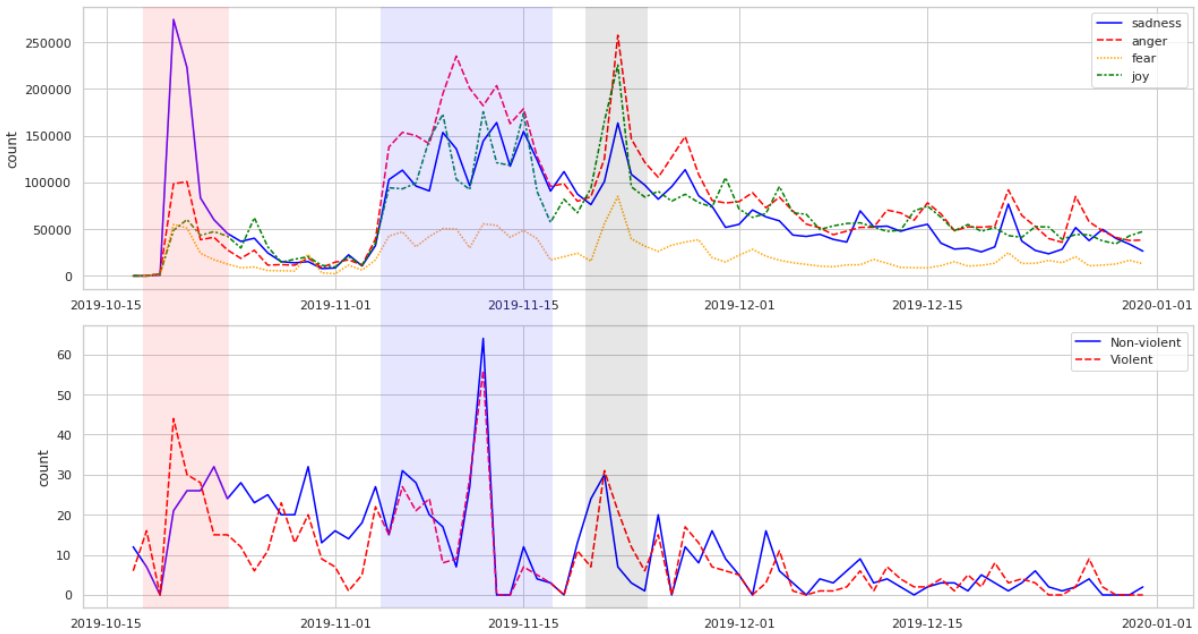

Figure 5.14: Top: Evolution of emotions in the Chilean 2019 Social Movement based on Twitter data. Bottom: Reported contentious violent (red) and non-violent (blue) events related to the Chilean 2019 Social Movement based on ACLED data. The transparent vertical bars of various colours represent important moments in the chronology of the movement.

The Chilean 2019 Social Movement has a sizable amount of reported contentious events (bottom half of the Figure 5.14) both violent (45% of the total events) and non-violent (54% of the total events). In terms of emotional patterns the Chilean 2019 Social Movement's most prevalent emotions is anger (34%) followed closely by joy (30%) and sadness (27%). This general emotion composition can be seen in Figures 5.14 in which the Chilean 2019 Social Movement maintains, in general, anger and sadness as its top emotions throughout the whole period of observation.

In Figure 5.14 I have highlighted three periods between October 2019 and December 2019 with three vertical transparent bars in red, blue and grey shades. The first highlighted period (red shade) corresponds to the start of the social unrest, the 17th to 18th of October of 2019. In Figure 5.14 this is marked by a considerable spike in sadness and in reports of violent events, followed by an increase in non-violent events. During those dates there were several massive demonstrations in Santiago de Chile mainly against police brutality displayed against secondary school students in the days previous to the demonstrations. This might offer an explanation for why the main initial emotional reaction is sadness, being mainly an expression of concern for the secondary school students.

The second highlighted period (blue shade) corresponds to the first serious attempt of the

Chilean government to forcefully contain the social unrest. On the 7th of November of 2019 President Piñera summoned the National Security Council (COSENA), an institution reserved for war times and created during the dictatorship era, to announce his new National Security Agenda, heavily aimed to suppress and criminalise any acts of protest. Naturally, this was received with anger by the Chilean 2019 Social Movement. In Figure 5.14 we can see that the blue highlighted period, starting with the 7th of November of 2019, shows a strong increase in anger that reaches its peak around the 10th to 12th of November. This is when the most massive and violent demonstrations took place in 2019, which is reflected in the spike of violent and non-violent reported events around 12th of November (see 5.14).

Finally, the third highlighted period (grey shade) starts on the 22nd of November of 2019, with the public release of an Amnesty International report denouncing systematic Human Rights violation perpetrated by the Chilean state against protesters. Piñera's government categorically rejected this accusation causing a mixture of emotional reactions among the social movement. As can be seen in Figure 5.14, the Chilean 2019 Social Movement Twitter data shows a spike of anger and joy after the declaration of Amnesty International. My interpretation of this is that the anger was mainly because of the verification (by a third party organisation) of the rumours of Human Rights violations, and the joy a reaction to the public international denunciation of the criminal actions of Piñera's government. Examples of tweets of this periods include (translated from Spanish): "The Chilean goverment's reaction to the Amnesty International report is frankly shameful."; "This government thinks stupidly that they have to only respond to the Chilean institutions. They are very wrong."; "The international community is starting to reject Sebastian Piñera, this is not trivial. An isolated government serves no one, much less the elite."

The reaction of Piñera to the Amnesty International report caused an outburst of social unrest, leading to the last big spike in reports of violent and non-violent contentious events during the tracked period (see bottom of Figure 5.14).

After these three highlighted periods the emotional display on Twitter and the reported contentious events of the Chilean movement started to steadily decline, mainly as a reaction to the "Agreement for peace" which started the process of rewriting the Chilean constitution, one of the main demands of the social movement.

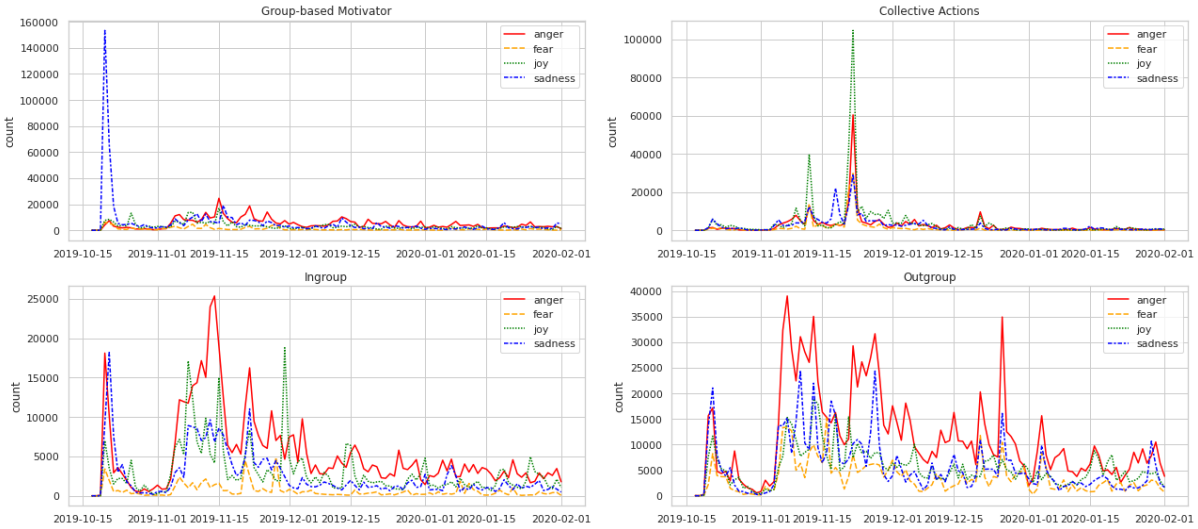

Figure 5.15: Evolution of emotions in the Chilean 2019 Social Movement per socio-psychological aspect of the movement.

Disaggregating the emotional evolution of the Chilean movement by each of their main aspects (Group-base Motivator, Collective Action, Ingroup and Outgroup) as shown in Figure 5.15, will help to see in greater detail how the social movement felt about the different aspects. This disaggregation will later be used to construct statistical models in order to test their predictive power with respect to violent and non-violent collective actions.

In the top-left corner of Figure 5.15 we can see the evolution of the Group-based Motivator emotions in the Chilean 2019 Social Movement In general the emotions were mainly sadness (38% of the total tweets) and anger (34% of the total tweets), showing a big spike of sadness at the very beginning of the measured period. The evolution of the emotions related to the Collective Actions is shown at the top-right corner of Figure 5.15. The general mood in collective action related tweets was joy (39% of the tweets) followed by anger (25% of the tweets). There are two important spikes of joy during the measured period, which correspond to the massive demonstrations on the 12th of November in response to Piñera's security agenda, and on the 22nd of November as a reaction to reports of human rights violations made by Amnesty International. The plot on the bottom-left corner of Figure 5.15 shows an unexpected finding: the most prevalent emotion towards the Ingroup was anger (46% of the tweets). The period in which the anger was most pronounced corresponds with the demonstration against Piñera's security agenda on November 13th, and the demonstrations in reaction to the reports of human rights violations, on November 22nd to 27th. Examples of angry tweets about the Ingroup include (translated from Spanish): "Hard to believe there are people supporting the protesters

on the streets. Just hate and violence"; "Here we have the protesters of the 'first line' looting and burning the Coquimbo Hospital." As it can be seen in these examples, it seems the emotion of anger towards the Ingroup is related to highly violent collective actions events, as a form of rejection of such ways to protests. These violent protest incidents were used by the opposition of the movement to blame the 'far-left' and attack the motivations and legitimacy of the demonstrations. We can also see that these angry tweets are mainly written from the perspective of someone not identified with the Ingroup, although the emotion is directed towards the Ingroup. This is sadly a limitation of my approach as there is no viable way to distinguish between the two by computational means.

Finally, the plot at the bottom-right corner of Figure 5.15 shows the emotional evolution of the tweets related to the Outgroup. In general terms, anger was the most prevalent emotion with 43% of the total tweets, which was a finding in line with the motivations of the social movement.

Statistical analysis

Table 5.20 shows the estimates of the VAR model using the trend of general emotions described in Figure 5.14 as predictors of Violent and Non-violent collective actions, for the Chilean 2019 Social Movement. Although a lag (order) of one period (day) was already set for all the models on the basis of theoretical reasons, Table 5.19 shows that three out of four lag selection criteria (Bayesian Information criterion (BIC) (Vrieze 2012), Akaike's Final Prediction Error (FPE) (Niedźwiecki and Ciołek 2017), Hannan–Quinn information criterion (HQIC) (Lopez and Weber 2017)) suggested an optimal lag length of one period (day) as well.

VAR order (lag)	AIC	BIC	FPE	HQIC
0	1.871	2.067	6.496	1.949
1	-1.305	0.06626*	0.2725*	-0.7614*
2	-1.315	1.231	0.2762	-0.3065
3	-1.316	2.405	0.2940	0.1586
4	-1.216	3.680	0.3680	0.7235
5	-1.171	4.900	0.4789	1.235
6	-1.017	6.229	0.7972	1.854
7	-1.831*	6.590	0.6216	1.506

Table 5.19: Optimal VAR order selection (lag) criteria for the Chilean 2019 Social Movement models (* highlights the best).

In Table 5.20 we can see that only violent, which events occurred at $t-1$ (the day before) have

	∇Non-violent events	Violent events	logAnger	logFear	∇ logJoy	logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.209(-1.72)	-0.018(-0.17)	-0.002(-0.45)	-0.002(-0.42)	0.002(0.35)	-0.003(-0.46)
$Violent_{t-1}$	-0.416(-2.76)*	0.295(2.17)*	0.004(0.64)	0.009(1.34)	0.005(0.78)	0.002(0.35)
$\log anger_{t-1}$	-1.767(-0.44)	-1.884(-0.52)	1.202(7.51)***	0.646(3.58)***	0.293(1.86)	0.553(2.91)***
$\log fear_{t-1}$	-1.583(-0.37)	3.139(0.82)	0.039(0.23)	0.275(1.44)	-0.172(-1.04)	0.051(0.26)
$\nabla \log joy_{t-1}$	3.172(1.33)	4.627(2.14)*	0.337(3.54)***	0.443(4.13)***	0.244(2.61)*	0.532(4.72)***
$\log sadness_{t-1}$	3.406(0.87)	-0.59(-0.17)	-0.587(-3.76)***	-0.411(-2.33)*	-0.434(-2.83)***	0.007(0.04)
No. of Equations:	6	BIC:	0.9955			
Nobs:	74	HQIC:	0.2095			
Log likelihood:	-576.4579	FPE:	0.7343			
AIC:	-0.3122					

Table 5.20: VAR model for Chilean 2019 Social Movement using general emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

a negative, statistically significant, effect on the de-trended reports of non-violent collective actions ($\nabla Non\text{-}violent$). On the other hand, the occurrence of violent collective actions the day before ($Violent_{t-1}$) and the detrended log of joy of the previous day ($\nabla \log joy_{t-1}$) have both statistically significant positive effects on the occurrence of violent collective actions. Finally, neither violent nor non-violent collective actions seem to have a predictive effect on any of the emotions, suggesting that the direction of the effects comes from the emotions towards the collective actions. These results suggest that, in the context of the Chilean 2019 social movement, violent collective actions have an inhibitory effect on the occurrence of non-violent collective actions the next day, but a boosting effect on the occurrence of violent collective actions. Additionally, general joy boosted the occurrence of violent collective actions. This interpretations are supported by the Granger causality test on these relationships, with violent actions causing a decrease in non-violent actions having a $Granger(1, 402) = 7.632, p = 0.006$, violent actions causing violent actions having a $Granger(1, 402) = 4.707, p = 0.031$, and joy causing violent actions having a $Granger(1, 402) = 4.589, p = 0.033$.

Tables from 5.21 to 5.24 show the results of the VAR models using the emotional trends of the Group-based Motivators, Collective Actions, Ingroup and Outgroup socio-psychological.

In the case of non-violent collective actions, **sadness related to the Outgroup** (see Table 5.24) was the only emotion that had a statistical significant effect on the occurrence of non-violent actions, showing positive effect with a $Granger(1, 390) = 4.509, p = 0.034$, suggesting this emotion would boost the occurrence of non-violent collective actions the next day. Examples of tweets containing sad messages mentioning the outgroup are (translated from Spanish): "Gustavo Gatica was blinded by Carabineros de Chile, Gustavo was blinded by the government of Piñera"; "Chile under the dictatorship of Piñera: 42 dead, 12 raped woman, 121 missing, thousands tortured." On the flip side, non-violent collective actions had a positive effect over the future feeling of **joy related to collective actions** ($Granger(1, 390) = 4.508, p = 0.034$, see Table 5.22), suggesting that the more reports of non-violent actions there are, the happier will be the conversation about collective actions the next day. Examples of joyful tweets mentioning collective actions are: "Santiago de Chile: beautiful picture of the protests"; "It's on: Calls for the biggest K-pop march this Friday in Santiago de Chile."

In the case of violent collective actions, **fear related to the Group-based Motivator** (e.g., "Director of Amnesty International Chile received death threats after releasing human rights

	∇Non-violent events	Violent events	logAnger	logFear	∇ logJoy	∇ logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.116(-0.9)	0.093(0.89)	0.004(0.72)	0.01(1.47)	0.006(0.94)	-0.002(-0.36)
$Violent_{t-1}$	-0.595(-3.13)***	0.18(1.16)	-0.003(-0.33)	0.009(0.82)	0.002(0.24)	0.01(0.98)
$\log anger_{t-1}$	-1.132(-0.7)	-0.272(-0.21)	0.829(11.61)***	0.003(0.03)	-0.003(-0.04)	0.035(0.4)
$\log fear_{t-1}$	3.372(1.65)	4.314(2.59)*	-0.025(-0.27)	0.666(5.96)***	-0.061(-0.56)	-0.209(-1.94)
$\nabla \log joy_{t-1}$	1.124(0.55)	-0.879(-0.53)	-0.096(-1.07)	-0.069(-0.62)	-0.158(-1.47)	0.166(1.54)
$\nabla \log sadness_{t-1}$	0.581(0.38)	0.516(0.42)	0.123(1.84)	0.039(0.47)	0.076(0.95)	-0.194(-2.42)*
No. of Equations:	6	BIC:	4.8172			
Nobs:	72	HQIC:	4.0178			
Log likelihood:	-696.5898	FPE:	32.8785			
AIC:	3.4891					

Table 5.21: VAR model for Chilean 2019 Social Movement using Group-based Motivator emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	logAnger	∇ logFear	logJoy	∇ logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.202(-1.59)	-0.009(-0.08)	0.013(1.36)	0.008(0.66)	0.021(2.12)*	0.001(0.08)
$Violent_{t-1}$	-0.205(-1.31)	0.5(3.73)***	0.001(0.12)	0.017(1.2)	0.018(1.48)	0.003(0.23)
$\log anger_{t-1}$	1.402(0.86)	-0.05(-0.04)	0.719(5.76)***	0.022(0.15)	0.279(2.18)*	-0.02(-0.15)
$\nabla \log fear_{t-1}$	-1.582(-1.03)	-0.691(-0.53)	-0.072(-0.62)	-0.352(-2.52)*	-0.086(-0.71)	0.195(1.53)
$\log joy_{t-1}$	-2.934(-1.57)	-0.438(-0.27)	0.112(0.78)	-0.229(-1.35)	0.444(3.04)***	-0.126(-0.81)
$\nabla \log sadness_{t-1}$	1.294(0.83)	0.475(0.36)	0.079(0.66)	0.143(1.01)	-0.058(-0.47)	-0.133(-1.03)
No. of Equations:	6	BIC:	7.4418			
Nobs:	72	HQIC:	6.6425			
Log likelihood:	-791.0768	FPE:	453.7087			
AIC:	6.1138					

Table 5.22: VAR model for Chilean 2019 Social Movement using Collective Actions emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	logAnger	logFear	logJoy	logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.152(-1.21)	0.019(0.18)	-0.003(-0.53)	0.002(0.19)	0.007(0.98)	0.004(0.87)
$Violent_{t-1}$	-0.396(-2.34)*	0.389(2.74)*	0.008(1.19)	0.021(1.89)	-0.003(-0.35)	0.013(2.01)*
$\log anger_{t-1}$	-0.372(-0.13)	-1.81(-0.73)	0.889(7.57)***	0.443(2.3)*	0.391(2.25)*	0.472(4.07)***
$\log fear_{t-1}$	-0.598(-0.28)	0.207(0.12)	-0.124(-1.48)	0.055(0.4)	0.092(0.74)	-0.018(-0.22)
$\log joy_{t-1}$	-2.58(-1.17)	-0.111(-0.06)	0.15(1.71)	0.187(1.3)	0.403(3.1)***	-0.075(-0.86)
$\log sadness_{t-1}$	4.013(1.47)	2.473(1.08)	-0.052(-0.47)	0.137(0.77)	-0.097(-0.6)	0.506(4.7)***
No. of Equations:	6	BIC:	4.3308			
Nobs:	72	HQIC:	3.5314			
Log likelihood:	-679.0792	FPE:	20.2147			
AIC:	3.0027					

Table 5.23: VAR model for Chilean 2019 Social Movement using Ingroup emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	Anger	Fear	Joy	Sadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.199(-1.69)	-0.04(-0.37)	0.002(0.33)	0.007(0.94)	0.0(0.0)	0.001(0.11)
$Violent_{t-1}$	-0.369(-2.76)*	0.431(3.57)***	0.001(0.07)	0.002(0.29)	0.005(0.83)	0.009(0.97)
$\log anger_{t-1}$	1.139(0.32)	0.355(0.11)	0.695(3.53)***	0.107(0.47)	0.433(2.65)*	0.146(0.62)
$\log fear_{t-1}$	-4.253(-1.61)	-4.05(-1.7)	0.032(0.22)	0.44(2.59)*	-0.082(-0.66)	0.257(1.46)
$\log joy_{t-1}$	-3.693(-1.12)	-4.094(-1.38)	-0.503(-2.72)*	-0.31(-1.46)	-0.194(-1.27)	-0.53(-2.41)*
$\log sadness_{t-1}$	5.7(2.12)*	5.38(2.22)*	0.344(2.29)*	0.375(2.18)*	0.367(2.94)***	0.684(3.82)***
No. of Equations:	6	BIC:	4.0748			
Nobs:	73	HQIC:	3.2822			
Log likelihood:	-680.1257	FPE:	15.8085			
AIC:	2.757					

Table 5.24: VAR model for Chilean 2019 Social Movement using Outgroup emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

violations report") ($Granger(1, 390) = 6.683, p = 0.010$, see Table 5.21) and **sadness related to the Outgroup** (e.g., "I am shocked. The path the government is taking is really dangerous. They have chosen pure and hard polarization, instead of talking and giving in. It is bad news for Chile, today is a sad day, due to the indolence of its government.") ($Granger(1, 396) = 4.941, p = 0.027$, see Table 5.24) have both positive statistically significant effects on the occurrence of violent collective actions, meaning that these emotions would boost the occurrence of the violent actions the next day. On the other hand, violent collective actions had a statistically significant positive effect over the feeling of **sadness towards the ingroup** (e.g., "What an immense shame... 'Between so much darkness and so much death, I gave away my eyes so that people would wake up' The heart of Chile bleeds from the eyes taken from Gustavo Gatica... Justice for him, whatever the cost!") with a $Granger(1, 396) = 4.055, p = 0.045$ (see Table 5.23) suggesting that the greater the reports of violent collective actions, the sadder the conversation about the ingroup is going to be the next day.

Finally, we can see again statistically significant positive effects in the models 5.21, 5.23 and 5.24 in which reports of violent collective actions at $t-1$ have a negative effect on the occurrence of non-violent collective actions, but positive statistically significant effects on violent collective actions (see Tables 5.22, 5.23 and 5.24).

5.4.4 Emotional patterns of the Hong Kong 2019 Social Movement

Overview

Figure 5.16 shows the evolution of the emotional reaction detected in the tweets of the Hong Kong 2019 Social Movement, as described in Table 4.2. The bottom half shows the evolution of reported contentious events around the 2019 Hong Kong protests.

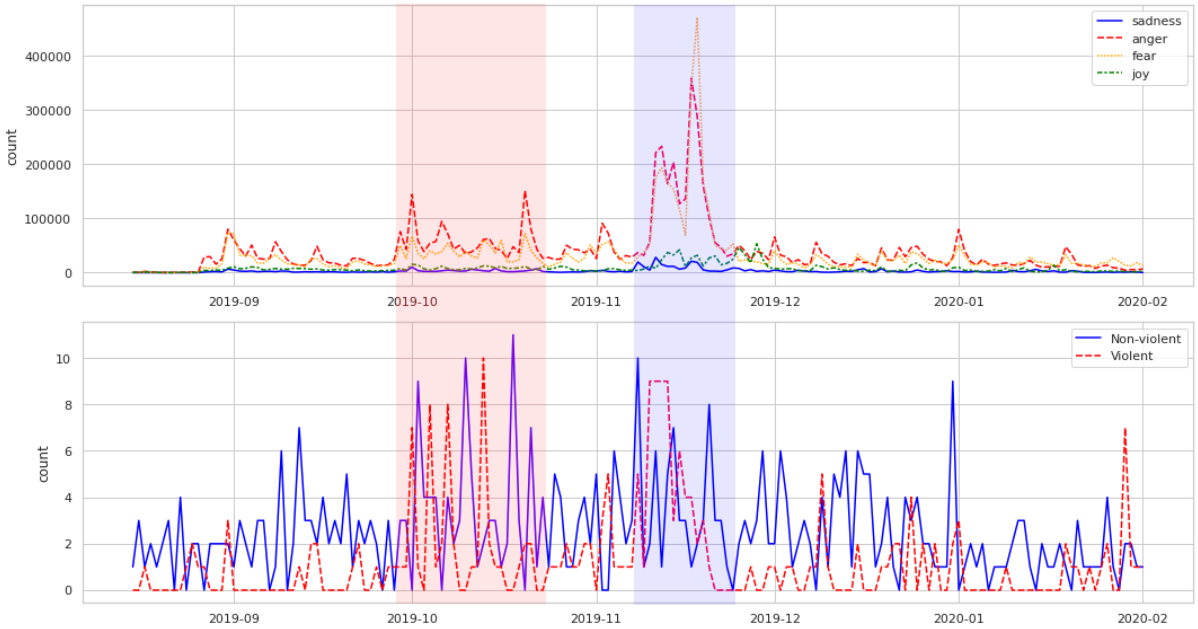

Figure 5.16: Top: Evolution of emotions in the Hong Kong 2019 Social Movement based on Twitter data. Bottom: Reported contentious violent (red) and non-violent (blue) events related to the Hong Kong 2019 Social Movement based on ACLED data. The two transparent vertical bars represent moments of special interest in the data.

In general terms, the most prevalent emotions in the tweets of the Hong Kong 2019 Social Movement are anger with 47% of the tweets showing this emotion, and fear with 39% of the tweets showing this emotion. In terms of the collective actions carried out by the movement, 70% of the reports correspond to non-violent events, while 30% correspond to violent events. The two highlighted periods in Figure 5.16 show moments of special interest, when emotional expressions were particularly pronounced. The first one, highlighted in a red shade in Figure 5.16 corresponds to the intensification of collective actions at the beginning of October 2019. The biggest spikes in emotional reactions are around the 1st of October 2019, dubbed the "National Day of Mourning" in protest against the celebrations of the 70th anniversary of the People's Republic of China, and around the 20th of October of 2019 when the Kowloon Protest took place against the government's decision to invoke the emergency law and in condemnation

of police brutality. However, during the whole month of October 2019 the number of contentions events was considerably high, and it was one the most intense months of the whole movement. The second period, highlighted in blue shade in Figure 5.16, started on the 11th of November of 2019 and ended on the 18th of November of 2019. This period is marked by an intense contentions events activity. Specifically, this is when the siege of the Chinese University of Hong Kong and the Hong Kong Polytechnic University happened, followed by violent storming of the besieged universities by the Hong Kong police, which ended with the arrest of several students. These events were one of the most notorious of the Hong Kong 2019 protest, and this was reflected in the high spikes of anger and fear during those dates. Examples of tweets from this period include: **Anger**: *"The Chinese and #HongKong governments must immediately deescalate the situation and exercise restraint at #PolyU."*; **Fear**: *"Resist now or we can never resist. Please support CUHK students and Hong Kongers. #HongKong"*.

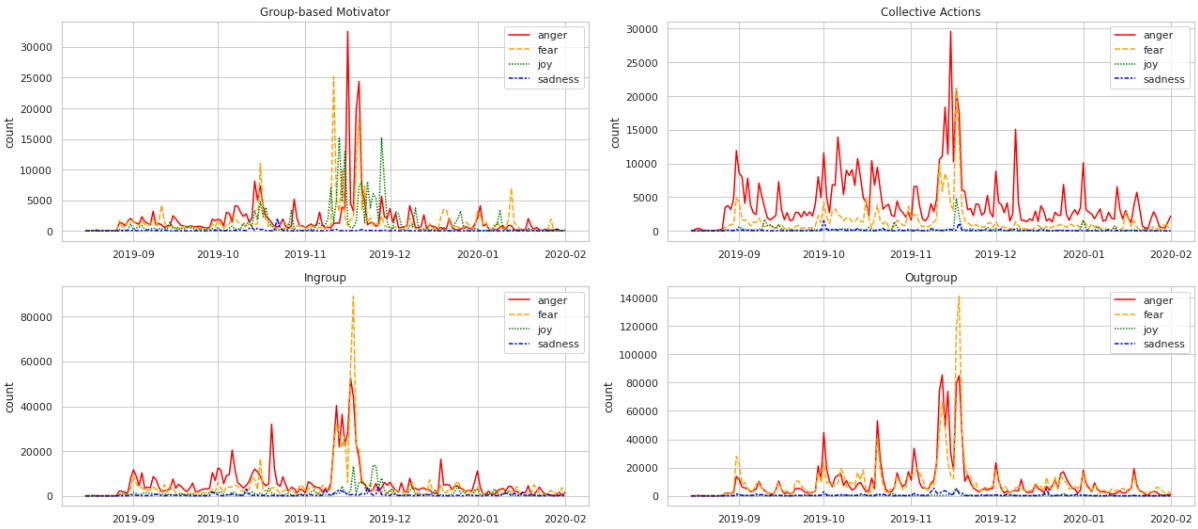

Figure 5.17: Evolution of emotions in the Hong Kong 2019 Social Movement per socio-psychological aspect of the movement.

At the top-left corner of Figure 5.17 we can see the evolution of the Group-based Motivator emotions of the Hong Kong 2019 Social Movement. We observe high levels of anger (40% of the tweets) and fear (31% of the tweets) but with the irruption of joy (27% of the tweets) at certain periods. There is an important spike of fear around 11th of November, the day of the start of siege of the Chinese Hong Kong University, followed by spikes of anger on the 13th and 15th of November, corresponding to days of generally peaceful demonstrations. This is followed by spikes of joy around the 23rd of November, one day before the Hong Kong Civil Councils elections, and around the 28th of November, when demonstrators celebrated the signing of the

Hong Kong Human Rights and Democracy Act by President Donald Trump. The plot at the top-right corner of Figure 5.17 shows the evolution of the emotions related to the Collective Actions of the Hong Kong 2019 Movement. In general, we can see that there are essentially only two emotions at play: anger, with 72% of the tweets, and fear with 23% of the tweets. These emotions show periods of great activity around the dates when massive and often violent collective actions took place, similar to the patterns seen in Figure 5.16. We see a period of high anger levels between the 1st and 20th of October of 2019, which is coincident with the intensification of the protest in the city of Hong Kong and a spike of anger related to the siege of Hong Kong Universities and its violent termination, between the 11th and 17th of November of 2019. We see the last significant spike of anger during the measured period on the 8th of December of 2019, which corresponds to the last massive protest in the year 2019 in Hong Kong.

Plots in the bottom-left and bottom-right of Figure 5.17 show the evolution of the emotions related to the Ingroup and Outgroup of the Hong Kong 2019 Social Movement, respectively. Both emotional patterns are dominated by anger (50% and 47% respectively) and fear (36% and 48% respectively). While tweets are generally angry when talking about the Ingroup, when talking about the Outgroup tweets are dominated generally by fear, with some anger. This is particularly evident in mid November 2019, which corresponds to the dates of the siege of the Hong Kong Universities and its violent breakup.

Statistical analysis

Table 5.26 shows the estimates of the VAR model using the trend of general emotions described in Figure 5.16 as predictors of violent and non-violent collective actions. As was the case with the Chilean 2019 VAR analysis, the lag (order) of the VAR models of the Hong Kong 2019 social movement was set to be of 1 period (day). Analysis of optimal VAR order (see Table 5.25) showed that two criteria scores supported one period as the optimal order (BIC and HQIC scores), while the other two criteria supported an optimal order of two periods (AIC and FPE).

VAR order (lag)	AIC	BIC	FPE	HQIC
0	-0.3915	-0.2777	0.6760	-0.3453
1	-3.197	-2.400*	0.04090	-2.873*
2	-3.310*	-1.830	0.03659*	-2.709
3	-3.271	-1.108	0.03820	-2.393
4	-3.029	-0.1822	0.04907	-1.873
5	-2.836	0.6945	0.06035	-1.402
6	-2.765	1.449	0.06612	-1.054
7	-2.773	2.124	0.06748	-0.7845

Table 5.25: Optimal VAR order selection (lag) criteria for the Hong Kong 2019 models (* highlights the best).

In Table 5.26 we can see that none of the emotions at $t-1$ had statistically significant effects on the occurrence of violent or non-violent collective actions. The bulk of the effect on these variables was taken by their auto-regressive terms at $t-1$. Non-violent actions were negatively predicted by the non-violent actions at $t-1$ with a $Granger(1, 972) = 53.91, p < 0.000$. This finding suggests that greater levels of reported non-violent collective actions inhibit the occurrence of more non-violent collective actions the next day. On the other hand, violent actions are positively predicted by violent actions at $t-1$ with a $Granger(1, 972) = 20.23, p < 0.000$, meaning that reports of violent collective actions boost the occurrence of violent actions the next day. This positive feedback loop effect was also present for the Chilean 2019 social movement (see Table 5.20). Finally, non-violent collective actions at $t-1$ had a small but statistically significant negative effect on the occurrence of violent actions the next day, with a $Granger(1, 972) = 6.462, p = 0.011$, meaning the reports of non-violent collective actions contribute to reduce the occurrence of violent actions the next day.

Tables from 5.27 to 5.30 show the results of the Hong Kong 2019 movement VAR models using the emotional trends of the Group-based Motivators, Collective Actions, Ingroup and Outgroup socio-psychological aspects.

The analysis shows again no statistically significant effects from any of the emotions on either violent nor non-violent collective actions. The auto-regressive effects of these two variables were again the main predictors in all of the VAR models shown in Tables 5.27 to 5.30. Non-violent collective actions at $t-1$ have again a negative statistically significant effect on itself at t, while violent collective actions at $t-1$ have, in all the models, a positive, statistically significant effect on itself at t, suggesting again that reports of non-violent actions inhibit the occurrence of non-

	∇Non-violent events	Violent events	logAnger	logFear	logJoy	logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non - violent_{t-1}$	-0.497(-7.34)***	-0.12(-2.54)*	-0.016(-1.09)	-0.009(-0.71)	-0.006(-0.45)	0.027(1.62)
$Violent_{t-1}$	0.091(0.82)	0.35(4.5)***	0.036(1.53)	0.026(1.18)	0.005(0.21)	-0.003(-0.11)
$\log anger_{t-1}$	0.696(1.2)	0.527(1.31)	0.589(4.82)***	0.115(1.01)	0.051(0.42)	0.065(0.45)
$\log fear_{t-1}$	-0.325(-0.59)	-0.352(-0.92)	0.097(0.83)	0.643(5.93)***	-0.005(-0.05)	0.262(1.9)
$\log joy_{t-1}$	-0.076(-0.2)	-0.189(-0.73)	0.105(1.35)	-0.028(-0.38)	0.568(7.35)***	0.097(1.05)
$\log sadness_{t-1}$	-0.22(-0.66)	0.193(0.83)	0.074(1.06)	0.14(2.14)*	0.195(2.82)***	0.521(6.29)***
No. of Equations:	6	BIC:	-2.2056			
Nobs:	169	HQIC:	-2.6678			
Log likelihood:	-1144.7003	FPE:	0.0506			
AIC:	-2.9835					

Table 5.26: VAR model for Hong Kong 2019 Social Movement using general emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	∇ logAnger	logFear	logJoy	∇ logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.482(-6.7)***	-0.136(-2.71)*	-0.032(-1.56)	-0.006(-0.27)	0.014(0.52)	0.031(1.06)
$Violent_{t-1}$	0.109(0.99)	0.393(5.09)***	0.067(2.12)*	0.061(1.91)	0.001(0.02)	-0.008(-0.17)
$\nabla \log anger_{t-1}$	0.087(0.35)	-0.125(-0.71)	-0.188(-2.6)*	-0.08(-1.1)	-0.05(-0.52)	-0.06(-0.59)
$\log fear_{t-1}$	0.128(0.53)	0.117(0.69)	-0.085(-1.23)	0.569(8.11)***	0.265(2.87)***	0.041(0.42)
$\log joy_{t-1}$	-0.092(-0.59)	0.035(0.32)	0.045(1.0)	0.098(2.15)*	0.671(11.27)***	-0.059(-0.94)
$\nabla \log sadness_{t-1}$	0.075(0.38)	-0.113(-0.83)	-0.041(-0.73)	-0.004(-0.08)	-0.074(-0.99)	-0.293(-3.71)***
No. of Equations:	6	BIC:	3.6638			
Nobs:	157	HQIC:	3.1783			
Log likelihood:	-1518.0704	FPE:	17.2291			
AIC:	2.8462					

Table 5.27: VAR model for Hong Kong 2019 Social Movement using Group-based Motivator emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	logAnger	logFear	logJoy	logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non - violent_{t-1}$	-0.487(-7.05)***	-0.122(-2.55)*	-0.02(-1.3)	-0.014(-0.85)	-0.0(-0.0)	-0.006(-0.26)
$Violent_{t-1}$	0.125(1.09)	0.318(4.0)***	0.008(0.3)	0.067(2.35)*	-0.042(-1.06)	-0.011(-0.27)
$\log anger_{t-1}$	-0.004(-0.01)	-0.044(-0.15)	0.373(4.05)***	0.163(1.56)	0.024(0.17)	0.144(0.97)
$\log fear_{t-1}$	0.099(0.29)	0.424(1.8)	0.251(3.38)***	0.605(7.2)***	0.207(1.75)	0.349(2.93)***
$\log joy_{t-1}$	0.107(0.48)	-0.246(-1.59)	0.042(0.85)	-0.026(-0.46)	0.489(6.29)***	0.011(0.13)
$\log sadness_{t-1}$	-0.141(-0.64)	0.003(0.02)	0.078(1.63)	0.056(1.02)	0.105(1.37)	0.377(4.88)***
No. of Equations:	6	BIC:	1.3301			
Nobs:	166	HQIC:	0.8623			
Log likelihood:	-1416.3064	FPE:	1.7212			
AIC:	0.5427					

Table 5.28: VAR model for Hong Kong 2019 Social Movement using Collective Actions emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	logAnger	logFear	∇ logJoy	∇ logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non-violent_{t-1}$	-0.484(-6.84)***	-0.131(-2.68)*	-0.02(-1.14)	-0.035(-1.9)	0.019(0.74)	0.005(0.17)
$Violent_{t-1}$	0.098(0.83)	0.368(4.57)***	0.079(2.66)*	0.089(2.96)***	0.091(2.13)*	0.085(1.78)
$\log anger_{t-1}$	0.047(0.16)	0.311(1.53)	0.607(8.16)***	0.136(1.78)	-0.051(-0.47)	0.094(0.78)
$\log fear_{t-1}$	0.068(0.24)	-0.127(-0.65)	0.053(0.74)	0.476(6.54)***	-0.256(-2.48)*	-0.387(-3.35)***
$\nabla \log joy_{t-1}$	-0.058(-0.27)	0.246(1.67)	0.086(1.6)	0.035(0.64)	-0.138(-1.77)	0.217(2.48)*
$\nabla \log sadness_{t-1}$	-0.003(-0.02)	-0.227(-1.79)	-0.112(-2.4)*	0.041(0.87)	-0.046(-0.68)	-0.302(-4.0)***
No. of Equations:	6	BIC:	2.5559			
Nobs:	160	HQIC:	2.0765			
Log likelihood:	-1460.077	FPE:	5.7491			
AIC:	1.7487					

Table 5.29: VAR model for Hong Kong 2019 Social Movement using Ingroup emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	∇Non-violent events	Violent events	logAnger	∇ logFear	∇ logJoy	logSadness
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$\nabla Non - violent_{t-1}$	-0.488(-6.86)***	-0.13(-2.64)*	-0.016(-0.93)	-0.004(-0.23)	0.006(0.26)	-0.006(-0.29)
$Violent_{t-1}$	0.11(0.91)	0.324(3.86)***	0.074(2.47)*	0.024(0.79)	-0.024(-0.62)	0.073(1.96)
$\log anger_{t-1}$	-0.161(-0.55)	0.188(0.93)	0.629(8.77)***	-0.186(-2.57)*	-0.081(-0.87)	0.143(1.6)
$\nabla \log fear_{t-1}$	-0.163(-0.47)	-0.2(-0.83)	0.081(0.95)	-0.036(-0.42)	-0.15(-1.34)	0.024(0.23)
$\nabla \log joy_{t-1}$	-0.07(-0.27)	-0.032(-0.18)	-0.012(-0.18)	0.016(0.24)	-0.058(-0.69)	0.04(0.5)
$\log sadness_{t-1}$	0.359(1.47)	0.256(1.52)	0.036(0.6)	-0.024(-0.4)	0.033(0.42)	0.521(6.94)***
No. of Equations:	6	BIC:	1.0158			
Nobs:	159	HQIC:	0.5344			
Log likelihood:	-1327.9801	FPE:	1.2282			
AIC:	0.2052					

Table 5.30: VAR model for Hong Kong 2019 Social Movement using Outgroup emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

violent actions the next day, while reports of violent collective actions boost the occurrence of these the next day. Additionally, the data shows that reports of violent actions, while boosting violent actions, inhibit the occurrence of non-violent collective actions the next day (see Tables 5.27 to 5.30).

Finally, reports of violent collective actions at $t-1$ had a positive, statistically significant effect on several emotions. Reports of violent actions had a boosting effect over: **anger related to the Group-based Motivator** (e.g., "Pro-Democracy #HongKong restaurant, Lung Mun Cafe in Hung Hom, was seen vandalised today," "We never ask for independence but #FiveDemandsNotOneLess . Don't use it as an excuse to cover our demands on freedom and democracy which are provided by the basic law. If you want HK as a window of China, let HK be HK. If we burn, u burn with us. #HKprotests") with a $Granger(1, 900) = 4.500, p = 0.034$ (see Table 5.27); **fear related to Collective Actions** (e.g., "We are fighting for our survival...if Hong Kong falls, the whole world falls. #HongKongProtests #EmergencyLaw #antiEALB #StandWithHongKong," "A youngster got shot by pepper-ball projectile in HIS EYE. #HKPoliceTerrorism #HongKongProtests #HongKong") with a $Granger(1, 954) = 5.509, p = 0.019$ (see Table 5.28); **anger, fear and joy related to the Ingroup** (e.g., anger: "Resist now or we can never resist. Please support CUHK students and Hong Kongers. #HongKong," fear: "Chinese University is currently under severe attack by #HongKongPoliceTerrorists with various ammunition. The CCP puppet has decided to kill ALL the young people in #HongKong #SOSHK #HKPoliceState," joy: "Free #HongKong ! For an independent Kong Kong, liberated from communist control, thriving, independent, young and free. The future belongs to millions like JoshuaWong, yearning to breathe free, the people everywhere aflame with the fire of freedom.") with a $Granger(1, 918) = 7.097, p = 0.008$, $Granger(1, 918) = 8.765, p = 0.003$ and $Granger(1, 918) = 4.519, p = 0.034$ respectively (see Table 5.29); and finally **anger related to the Outgroup** (e.g., "This is vile. Communist China truly has no shame. The pro-democracy protestors in #HongKong are simply fighting for freedom," "Police has been firing tear gas into the City University of Hong Kong. The communist party wants to destroy our future") with a $Granger(1, 912) = 6.112, p = 0.014$ (see Table 5.30).

5.4.5 Emotional patterns of Fridays for Future

Overview

Figure 5.18 shows the evolution of the emotional content in the tweets for the Fridays for Future social movement, using the dataset of tweets described in Table 4.2.

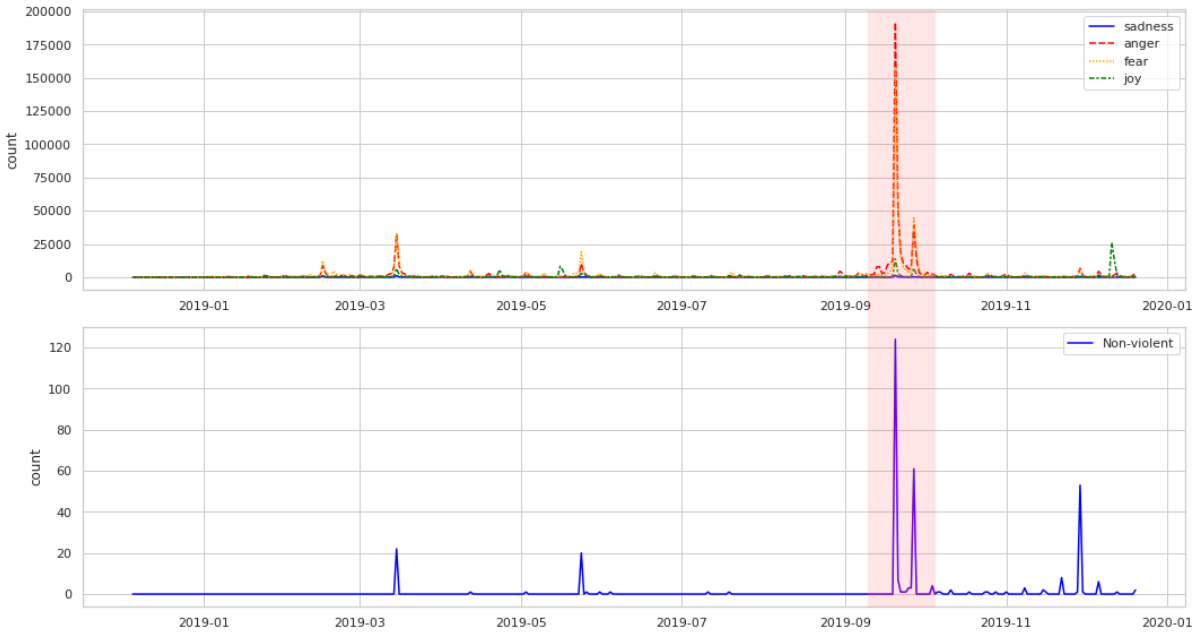

Figure 5.18: Top: Evolution of Fridays for Future (FFF) Social Movement emotions based on Twitter data. Bottom: Reported contentious non-violent (blue) events based on ACLED data. The transparent vertical red bar highlights the moment of greater activity.

From Figure 5.18 we can extract that the movement's communications on social media has anger and fear as their main emotions. Anger is present in 45% of the tweets while fear is present in 43% of them. We can also see that the emotional activity is concentrated in very few, relatively short, periods of time. The spikes occurred around the dates of the global school strikes events, 15th March, 25th May, 20th and 27th of September and 29th November. Emotions are particularly strong within the period highlighted in red shade in Figure 5.18, between the 20th and 27th of September of 2019, when the "Global Week for Climate Action" took place, which involved more than 2 million people in 150 locations worldwide. We also see a small spike in joy emotions in December during the 2019 United National Climate Change Conference (COP25) in Madrid, where FFF activists received a lot of attention and appreciation.

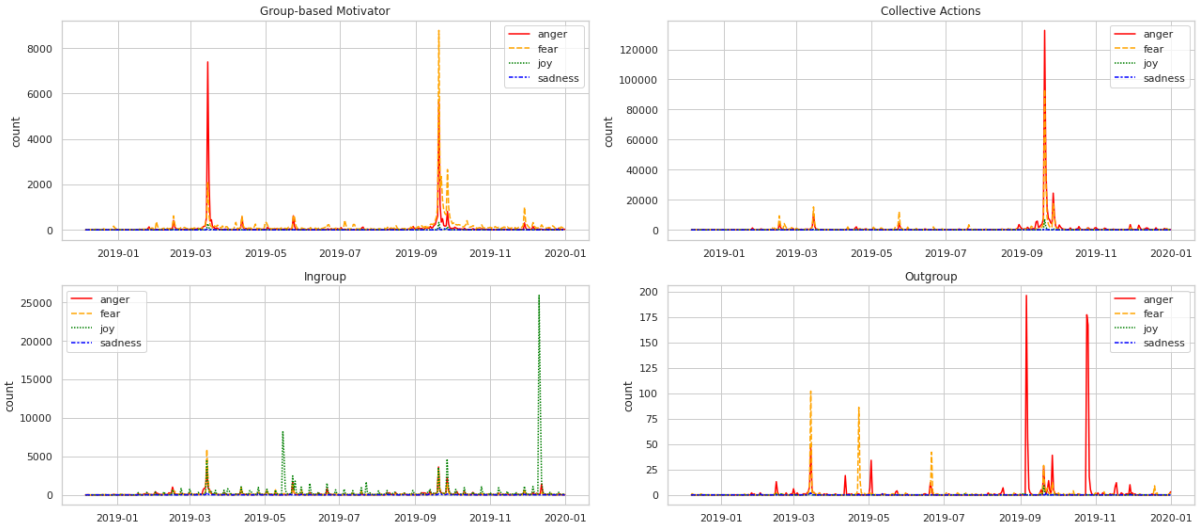

Figure 5.19: Evolution of emotions in the Fridays for Future movement, per socio-psychological aspect of the movement.

On the top-left corner of Figure 5.19 we see the plot of the emotions related to the Group-based Motivator of Fridays for Future. The tweets here are dominated by the emotions of fear (63%) and anger (33%). The top-right plot of Figure 5.19 shows the evolution of the emotions related to the Collective Actions of Fridays for Future. Here, anger accounts for 54% of the tweets, followed by fear with 44% of the tweets. The bottom-left plot of Figure 5.19 shows the evolution of the Ingroup related emotions. Here, the main emotion is joy, with 58% of tweets, driven mainly by a big spike in joy on the 11th of December of 2019, corresponding to the date when the Time magazine named Greta Thunberg, original initiator of the Fridays for Future movement, Person of the Year. This announcement was made while Greta was participating in the COP25, in which she and Friday for Future had a major role as an activist organisation. Other emotions related to the Ingroup are anger (21%) and fear (19%). Finally, the bottom-right plot in Figure 5.19 shows the evolution of the emotions related to the Outgroup. I must note that the number of tweets in the Fridays For Future data referring to the Outgroup was relatively small (1631 tweets) and hence it is difficult to draw a reliable picture of the temporal evolution of the emotions with respect to the Outgroup. Nevertheless, the collected tweets referring to the Outgroup were dominated by anger (49%) and fear (47%), with periods of great activity around the 15th March of 2019 and 20th and 27th of September of 2019, when the Global Strike for Climate took place.

Statistical Analysis

Table 5.32 shows the estimates of the VAR model using the trend of general emotions described in Figure 5.18 as predictors of the Non-violent collective actions of the Fridays for Future movement. Analysis of optimal VAR order can be seen in Table 5.31). All optimal VAR order criteria supported one period as the optimal lag.

VAR order (lag)	AIC	BIC	FPE	HQIC
0	2.723	2.780	15.23	2.746
1	1.162*	1.505*	3.195*	1.298*
2	1.218	1.847	3.381	1.469
3	1.268	2.183	3.557	1.633
4	1.355	2.556	3.880	1.834
5	1.415	2.902	4.124	2.008
6	1.300	3.072	3.678	2.006
7	1.234	3.293	3.450	2.055

Table 5.31: Optimal VAR order selection (lag) criteria for the Fridays for Future models (* highlights the best).

Table 5.32 shows that there are no statistically significant effects of any general emotions at $t-1$ on the occurrence of non-violent collective actions of the Fridays for Future movement. Coincidentally, non-violent collective actions at $t-1$ had no statistically significant effect on any of the emotions expressed in the tweets the next day.

Tables from 5.33 to 5.35 show the results of the Fridays for Future movement VAR models using the emotional trends of the Group-based Motivators, Collective Actions and Ingroup socio-psychological aspects. The VAR model using the emotional trends with respect to the Outgroup was not possible to be estimated due to the small number of observations (days) with at least one tweet referring to the Outgroup.

The results show that none of the emotions at $t-1$ have a statistically significant effect over the non-violent collective actions, and that the reports of non-violent collective actions at $t-1$ had no statistically significant effect on any of the emotions in any of the estimated models. Likewise, no statistically significant auto-regressive effects on violent or non-violent collective actions were found in any of the models.

	Non-violent events	logAnger	logFear	logJoy	logSadness	
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$Non-violent_{t-1}$	-0.06(-1.02)	-0.003(-0.6)	0.001(0.12)	-0.001(-0.1)	-0.002(-0.32)	
$\log anger_{t-1}$	1.186(1.66)	0.907(13.29)***	0.297(4.31)***	0.344(3.56)***	0.353(3.89)***	
$\log fear_{t-1}$	0.749(0.96)	-0.022(-0.29)	0.509(6.8)***	-0.129(-1.22)	-0.01(-0.1)	
$\log joy_{t-1}$	-0.824(-1.71)	-0.102(-2.22)*	-0.081(-1.75)	0.426(6.57)***	-0.023(-0.37)	
$\log sadness_{t-1}$	0.48(0.89)	-0.036(-0.7)	-0.037(-0.72)	-0.045(-0.62)	0.309(4.5)***	
No. of Equations:	5	BIC:	1.4994			
Nobs:	339	HQIC:	1.2957			
Log likelihood:	-2571.8561	FPE:	3.1925			
AIC:	1.1608					

Table 5.32: VAR model for Fridays for Future using general emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	Non-violent events	Anger	Fear	Joy	Sadness	
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$Non-violent_{t-1}$	-0.145(-1.72)	-0.007(-0.88)	-0.006(-0.99)	-0.001(-0.07)	-0.011(-1.87)	
$\log anger_{t-1}$	-0.054(-0.06)	0.589(6.84)***	0.103(1.79)	0.123(1.57)	0.045(0.74)	
$\log fear_{t-1}$	2.325(1.79)	0.209(1.63)	0.507(5.92)***	0.147(1.25)	0.281(3.05)***	
$\log joy_{t-1}$	0.904(0.92)	-0.062(-0.63)	0.029(0.45)	0.146(1.65)	-0.058(-0.84)	
$\log sadness_{t-1}$	1.043(0.96)	0.111(1.03)	0.121(1.7)	0.146(1.5)	0.39(5.1)***	
No. of Equations:	5	BIC:	2.8197			
Nobs:	189	HQIC:	2.5136			
Log likelihood:	-1528.7325	FPE:	10.0266			
AIC:	2.3051					

Table 5.33: VAR model for Fridays for Future using Group-based Motivator emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	Non-violent events	Anger	Fear	Joy	Sadness	
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$Non-violent_{t-1}$	-0.137(-1.58)	-0.002(-0.31)	-0.002(-0.23)	-0.007(-0.73)	-0.004(-0.56)	
$\log anger_{t-1}$	0.682(0.65)	0.621(6.71)***	0.035(0.35)	0.15(1.37)	0.131(1.49)	
$\log fear_{t-1}$	0.164(0.16)	-0.04(-0.44)	0.509(5.13)***	0.227(2.13)*	0.127(1.49)	
$\log joy_{t-1}$	0.364(0.4)	0.177(2.21)*	0.145(1.65)	0.457(4.84)***	0.12(1.58)	
$\log sadness_{t-1}$	1.629(1.53)	-0.03(-0.33)	0.06(0.58)	0.017(0.16)	0.343(3.87)***	
No. of Equations:	5	BIC:	3.6742			
Nobs:	178	HQIC:	3.3554			
Log likelihood:	-1512.1334	FPE:	23.0596			
AIC:	3.138					

Table 5.34: VAR model for Fridays for Future using Collective Actions emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

	Non-violent events	Anger	Fear	Joy	Sadness	
	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)	Est.(t)
$Non-violent_{t-1}$	-0.013(-0.22)	0.002(0.25)	0.007(0.99)	0.006(0.77)	0.009(1.1)	
$\log anger_{t-1}$	0.847(1.31)	0.508(6.46)***	0.091(1.22)	-0.025(-0.29)	0.042(0.49)	
$\log fear_{t-1}$	0.429(0.57)	0.074(0.81)	0.385(4.41)***	0.084(0.83)	0.09(0.91)	
$\log joy_{t-1}$	-0.42(-0.77)	-0.007(-0.11)	-0.001(-0.01)	0.45(6.13)***	-0.0(-0.0)	
$\log sadness_{t-1}$	0.169(0.3)	-0.059(-0.87)	-0.03(-0.47)	-0.153(-2.05)*	0.221(3.0)***	
No. of Equations:	5	BIC:	3.4213			
Nobs:	299	HQIC:	3.1986			
Log likelihood:	-2547.2916	FPE:	21.1164			
AIC:	3.05					

Table 5.35: VAR model for Fridays for Future using Ingroup emotions. Signif. codes: *** < 0.001, ** < 0.01, * < 0.05.

5.5 Discussion: Emotions and Collective Action

5.5.1 Emotional patterns of social movements

In Section 5.4 I presented the evolution of four emotions (anger, fear, sadness and joy) for three social movements: Chilean 2019 Social Movement, Hong Kong 2019 Social Movement and Fridays for Future, based on Twitter data. The evolution of the four emotions was presented both in general terms, using the full data per movement as detailed in Table 4.2, and separated by socio-psychological aspect, proposed in Chapter 3, using the data detailed in Table 5.16.

The first major finding of this thesis is that **anger** is the most prevalent emotion among the three studied social movements. This emotion is present in 34% of the tweets of the Chilean 2019 movement, 47% of the tweets of the Hong Kong 2019 movement and 45% of the tweets of the Fridays for Future movement. This finding is consistent with the current literature in social psychology about social movements and collective action participation (Van Zomeren, Leach, et al. 2012; Drury and Reicher 2009; Tausch, Becker, Spears, et al. 2011b; Tausch and Becker 2013), in which anger is frequently characterised as one of the main factors predicting the participation in collective actions as well as a integral part of the motivational process involved in the formation of social movements.

The anger in the tweets was generally related to the specific struggle the social movement is fighting against: Police repression and oppressive government, in the cases of the Chilean and Hong Kong 2019 social movements; or climate change and the inaction of the world governments and society about it, in the case of the Fridays for Future movement, which is again consistent with the motivational models of collective action participation in social and political psychology (Van Zomeren, Leach, et al. 2012; Drury and Reicher 2009; Tausch, Becker, Spears, et al. 2011a) in which anger is a response to a group-wide grievance affecting the participants of the social movement.

However, the statistical analysis using VAR models revealed no statistically significant relations between the **general** trends of **anger** and the daily reports of collective action in any of the studied social movements. This means that, although anger is the main expressed emotion in the tweets, its general daily variations does not predict the reports of collective action, neither violent nor non-violent, the next day (see Tables 5.20, 5.26, 5.32). The opposite effect is also absent. There is no statistically significant effect of violent or non-violent reported actions on

the general anger trend (see Tables 5.20, 5.26, 5.32).

When splitting the data into aspects (see sections 2.5 and 5.3), analysis reveal that in the case of the Hong Kong 2019 movement, reports of violent collective actions have a positive statistically significant effect on anger related to the Outgroup, Ingroup and the Group-based motivator (see Tables 5.30, 5.29, 5.27). Even if the tweets are in general speaking about one of the aspects of the movement, the emotion of anger seems to be always associated to the initial struggle of the lack of democracy and civil rights and the abuse of power of the police forces, suggesting that the observed anger produced by the violent demonstrations is a reaction to the behaviour of the police and the Hong Kong government towards the ingroup and the ingroup demands, which is again consistent with the current motivational models in collective action literature, which claim that the anger must originate in a group-wide grievance.

Contrary to mainstream social and political psychology models, however, **anger** didn't statistically predict violent or non-violent collective actions in any of the tested models, which is an unexpected result. Moreover, the effects of anger, although not statistically significant, were inconsistent across the models, showing positive and negative effects alike. This suggests that the mechanism through which anger affects the willingness to participate in collective actions is more complex than previously hypothesised, possibly not being the proximal predictor of collective action, but maybe a distal predictor mediated by a non-observed variable.

However, these results are not intended to question the validity of any of the current mainstream social psychological models of collective action, but rather to highlight the need to test them in contexts different from the ones they were developed. Reicher (1996), Drury and Reicher (2009), and Van Zomeren, Leach, et al. (2012) models were developed using, mainly, self-reported data (interviews or questionnaires) and observant participation methods. These methods are very rich in subject-centred descriptions of the motivational process of the participants of social movements. Using such data, conventional models that describe neatly the influence of emotions on the participation in collective actions make complete sense, as they are designed to describe the general process of motivation. When these models are applied to time-series data in the expectation to produce a precise prediction of a future behaviour, however, many other contextual factors might be influencing the contingent decision of the participant. Moreover, the VAR models used in this analysis were estimated using intentionally strict hyper-parameters. In particular, the lag (order) of the models was constrained to be one

day, limiting the models to test the effects of anger at $t-1$ (a day before) on the reported collective actions of the current day. These constraints were justified in the case of my analysis in which I was specifically looking for the immediate effect of the emotions on collective actions, but it is perfectly reasonable to formulate hypotheses in which emotions would have a more delayed effect on the collective actions. The fact that anger is still the main emotion expressed among the social movements communications is still, in my opinion, evidence supporting the psychological models of collective actions.

The second major finding of this thesis is the consistent feedback loop in many of the models of the reported collective actions at $t-1$ on the reported collective actions of the current day. Specifically, results show that reports of non-violent collective actions have a negative inhibitor effect on the occurrence of violent and non-violent collective actions the next day, while reports of violent collective actions boost the occurrence of violent collective actions, while inhibiting the occurrence of non-violent collective actions in the future (e.g., Tables 5.20 5.26). These effects were found to be significant in the Chilean 2019 and Hong Kong 2019 social movements data. This feedback loop has been previously discussed in the literature on collective actions (Van Zomeren, Leach, et al. 2012), but the differential effects seen here are new. This finding could potentially have interesting policy implications. The fact that reporting the occurrence of non-violent collective actions leads to the decrease of violent action, and reports of violent action lead to more violent actions is an insight that could contribute to more responsible news reporting about violence and collective actions. Similar effects exist in the field of suicide prevention (Phillips 1974), where suicide reporting guidelines have been established to prevent copycats suicides. Another, equally valid interpretation of this finding is that violent collective actions could simply last more than one day, which in turn would lead to the results we are seeing in the analysis. However, this was not generally the case neither in the Chilean 2019 nor Hong Kong 2019 social movements, with violent demonstrations only lasting several hours. Alternatively, this could also be an artefact of the news sources the event data were collected from, in which news outlets continue to report events after the day they took place. Whatever is the case, more research is needed to test the robustness of the effects found in this thesis.

Third, the analysis revealed that besides anger other emotions are also present in the social media communications of social movements. **Fear** was the second most important emotion for the Hong Kong 2019 and Fridays for Future movements, with 39% and 43% of tweets respectively

showing this emotion.

Fear is known to be a powerful inhibitor of collective action participation (Miller et al. 2009) since, as was discussed in Chapter 2, it triggers a primitive neural circuit involved in the avoidance of danger. However, in none of the VAR models using data from the Hong Kong 2019 movement or the Fridays for Future movement fear had any statistically significant effect on the reports of collective actions the next day. Moreover, fear has only a statistically significant effect predicting positively reports of violent collective action in the case when this fear was related to the group-based motivator in the data from the Chilean 2019 social movement (see Table 5.21). Although this finding might seem contradictory, Choma et al. (2020) and Shepherd et al. (2018) shows that fear can also be a predictor of collective actions if it is in "fear-based threat" form, leading participants to engage in confrontational collective actions. The reason this effect is not replicated in the data of the Hong Kong 2019 social movement is unknown, although a possible explanation could be found in the level of perceived self-efficacy that movement have. In the tweets of the Hong Kong movement, fear is more related to the physical integrity of the protesters and the attacks of the police during the demonstrations, which could be understood as an indicator of low perceived self-efficacy. In this situation, it is expected that fear is not positively related to collective actions as the communications highlight the dangers of actions.

Fourth, **joy** only had a statistically significant effects on violent collective actions in the case of the Chilean 2019 social movement (see Table 5.20).

Self-directed positive emotions, such as joy, can have an effect on the participation in collective action. This hypothesis was tested before by Becker, Tausch, and Wagner (2011a). In their study, however, they did not find any evidence that positive emotions could boost participation in collective actions. Here I found evidence that joy can actually predict violent collective actions but also that non-violent collective action can Granger cause joy related to collective actions in the Chilean 2019 social movement data (see Table 5.22), further supporting the idea set by Tropp and Brown (2004) that collective action could cause self-affirming emotions in the participants. These findings should be understood in the context of the feedback loop that reports of violent and non-violent have on themselves. The inhibiting effect of reports of non-violent collective actions on future violent and non-violent collective actions could be mediated by increased joy related to collective actions. Nevertheless, further research combining self-reported data of the participants with third party reports of collective actions is needed to

determine precisely the details of this mechanism.

Fifth, **sadness** related to the outgroup (Chilean 2019 social movement, see Table 5.24) showed to have a positive, statistically significant effect on non-violent and violent collective actions.

The emotion of sadness is not usually considered as a motivator of collective action nor as a modulator of the activism itself. It is not until recently that sadness was considered as a possible motivator of participation in collective actions in the context of supporting disadvantaged groups (Lantos et al. 2020) as a feeling of "pity" for the disadvantaged, and in the context of supporting climate action (Landmann and Rohmann 2020) as a feeling of "being moved" to participate in activism. These results add to the relatively recent literature on how the emotion of sadness can intervene in the motivation process of collective actions. However, more work is needed to determine which are the conditions in which sadness can have a substantive effect.

Finally, the results suggest that different emotions towards different targets do play a role in the dynamics of collective actions, both violent and non-violent. However, the results vary dramatically between social movements, with emotions that are very important for one being totally absent in another. This diversity of results highlights, in my opinion, the importance of context in the study of social movements and their collective actions. The Hong Kong 2019 social movement, for example, faced a very strong and potentially militarised opposition, which explains in part the high levels of fear in their communications. In the case of the Chilean 2019 social movement on the other hand, the prospect of an intervention by the military in the contention of the social movement, as was the case during the Pinochet dictatorship, was seen as less likely, possibly influencing the low levels of fear present in their tweets. Sadly, using data of only three social movements it is impossible to reliably determine how these contextual variables might influence the expression of emotions. More research including a larger sample of social movements in a wider variety of contextual settings would be needed for such a purpose. Nevertheless, I believe this work at least highlights the importance of considering other emotions, besides group-based anger, and a wide variety of targets of those emotions, in the study of social movement and violent and non-violent collective actions.

5.5.2 Methodological reflections

VAR models were used to estimate the relationship between the time-series representing the emotions and the time-series representing the reports of violent and non-violent events. VAR

models are one of the standard techniques used in the modelling of multivariate time-series analysis (Toda 1991; Toda and Phillips 1994; Abrigo and Love 2016; Freeman et al. 1989) and it is widely used in econometrics, sociology and political science. However, the method is very sensitive to hyper-parameter changes. Specifically, the order of the model (lag) has a big impact on the viability of the model, often making the difference between significant and non-significant results. Moreover, the literature on VAR models seems to suggest that order selection (lag) is an optimisable problem (Ivanov and Kilian 2005), meaning that it is possible to estimate the order of the model that will deliver the best fit to the data. While this practice might be justifiable in the context of creating optimal predictive models (usually used in forecasting) it is less desirable in the context of hypothesis testing or theoretically guided exploration. Bruns and Stern (2019) suggests that optimal lag selection in VAR models could lead to undesirable research practices like "p-hacking" and "HARKing." P-hacking has been defined as the practice of tuning the data of the hyper-parameters of the models until statistically significant results appear. This practice is very prevalent in many disciplines of science and is boosted by the problematic incentives of "positive results" in publishing (Head et al. 2015). HARKing on the other hand, is the action of Hypothesising After the Results are Known and it is a practice designed to artificially inflate "positive results" in hypothesis testing (Kerr 1998). Given this, in this thesis I decided to set the lag a priori in order to restrict the range of the hypotheses and prevent harking and p-hacking. This, naturally, severely constrains the model and leads to a lower probability of statistically significant results, but provides clearer and more interpretable conclusions.

Under the conditions exposed above, and despite my best efforts on clarity, transparency and justification of methodological decisions, it is undeniable that a non-negligible probability exists that the outcomes presented in this work could be the results of the particular choices made by me in the process of analysis, instead of real patterns encountered in the data. Simonsohn et al. (2020) proposed a solution to this problem using what he named "specification curved analysis." This technique essentially proposes that the same model should be tested on the largest array of possible conditions in order to quantify its robustness. Future research will consider the development of a specific version of "specification curve analysis" for VAR models in order to test robustly the effects of emotions over collective actions.

5.5.3 Limitations of emotion analysis on Twitter and News Reports data

Among the most important limitations of this thesis is the low number of social movements studied. This limitation does not allow for statistical inference at the "movement level," constraining also the inferences I can make about the effects of contextual factors in which each social movement is immersed. To correct this limitation the collection of tweets from a large quantity of social movements or the collection of tweets related to a large number of collective actions both violent and non-violent, is necessary. That endeavour would require resources far beyond the scope of this project and is better suited for future, longer term work.

Additionally, the identification of the socio-psychological aspects by means of topics discovery analysis, while rich, leaves the door open for a fair amount of researcher bias. The keywords and aspects I deemed worthy of identification are not exhaustive, neither they pretend to be. The development of automatic methods of "most relevant" topic detection, based on a standard numerical criteria, would help reduce researcher bias. Future work should be aimed at exploring the development of such methods.

On the subject of data diversity and data quality, the use of more diverse sources of social media data instead of relying only on Twitter will be the next step. While Twitter is an excellent source of social media data, relaying on only one platform introduces possible sampling and user bias in the analysis. The use of social media data from other platforms, however would not have solved completely the platform bias problem, but it would have provided at least a point of comparison between each platform.

It is also important to note that social norms could be influencing the expression of emotions on Twitter. Specifically, expression of certain emotions and when they are being expressed could be guided by the social norms of the studied movements and the social norms of the societies they are embedded in. E.g., in societies where strong expressions of anger are frowned upon, we could expect less expression of anger in favour of the expression of other emotions (Hareli et al. 2015). Likewise, the participation in collective action is also mediated by the social norms of the social movement and the society. E.g., social movements with strong social norm of peaceful activism will exhibit less violent collective action (if any), while social movements with a social norm of participation in highly disruptive collective actions will exhibit more disruptive activism (for an example see Smith, González, et al. (2021)).

Social norms are a complex phenomena and they are worthy of their own study. The focus here was the accurate detection of a very limited set of emotions and to what extent these emotions can be related to violent and non-violent collective actions. Future work could take these findings and use the social norm framework to explore more nuanced ways to understand the relationship of emotions and collective actions. As an example, recent research by Spaiser, Nisbett, et al. (2022) has already used Twitter data to study social norms in the context of climate activism to explore how normative change can lead eventually to a wider societal change.

Finally, reports of events data coming from the ACLED dataset, although excellently curated, does rely on reports of news media in order to quantify the amount of contentious events, which might leave out smaller, but equally important events occurring in less known locations. Additionally, relying only on news reports introduces the bias of the news companies on what is marketable to report. Peaceful, low profile collective actions could be underrepresented in this kind of data, as their actions might not be disruptive enough to make the newspapers. Future work should diversify the sources of events data, in order to reduce these sampling biases.

Chapter 6

Conclusion

At its core, this thesis was an exploration, from the perspective of a social scientist, of the capacity that digital traces and social media data have to provide meaningful results in the study of social movements. As such, I tried to explore in depth both the methods through which social scientists can make sense of social media data as well as the data itself, trying to provide a thorough explanation of the processes through which the data was obtained as well as explorations of the inner workings of each of the techniques and algorithms used throughout the analysis. In this concluding chapter I will provide a brief summary of the results as well as an outline of what I consider to be the main contributions of this thesis.

6.1 Summary and contributions

In this thesis I set myself to explore the relationship between emotions expressed tweets and reported event of collective actions, guided by the most recent socio-psychological models on collective actions, to test at least the following hypotheses (see Section 2.5):

1. Anger will generally be positively associated with non-violent collective actions. Specifically, and following the evidence provided by the socio-psychological literature of collective actions, anger towards the outgroup and/or anger towards the group-based motivator will be positively associated with non-violent collective actions.

2. Anger will generally be negatively associated with violent collective actions. Specifically, anger towards the ingroup and/or anger towards the collective actions will be negatively associated with violent collective actions.

3. Fear would generally be negatively associated with violent and non-violent collective actions. Specifically, fear towards the collective actions and fear towards the outgroup would be negatively associated with any type of collective actions.

To my surprise, none of them found support in the analyses, revealing instead a much more complex panorama, in which emotions behave very differently depending on which aspect they were referring to, on which social movements were associated, and on which context they were measured. Despite this apparent defeat, the goal itself of attempting to measure emotions on tweets to later estimate their relationship with reports of real life contentious events, set me on a path of very interesting and fruitful challenges.

Giving that the detection of emotions in text was one of the core tasks in this thesis, one of the first challenges to tackle was to determine which emotions to measure. As it was outlined in Chapter 2, different scholars in different fields have come up with different taxonomies of basic emotions, many of which are based in rather outdated cross-cultural studies to justify their universal nature. I argued that there is already enough neuro-psychological evidence to settle on a relatively reduced number of basic emotions: Fear, Anger, Joy and Sadness. This simple taxonomy allowed me to have minimum overlap between the emotions, helping considerably in the process of emotions identification. Further in Chapter 2 I described the motivational process and the current socio-psychological models of collective action, outlining that the majority only consider one emotion (group-based anger) as a predictor for collective actions, highlighting the need for more research in the role of other basic emotions. There, I proposed a very simple model of the influence of the four basic emotions in violent and non-violent collective actions, which introduces a taxonomy of four basic internal "aspects" of social movements (see Section 2.5). The taxonomy includes the familiar **Ingroup** and **Outgroup** categories, usually conceptualised in socio-psychological model as the main targets of emotions, but adds the **Group-based Motivator** and **Collective Actions** of the social movement as also targets of emotions, drawing from the work of Tausch and Becker (2013) and Tausch, Becker, Spears, et al. (2011a).

Chapter 4 starts with an in-depth description of the data collection procedure, as well as the challenges of collecting large amounts of text data. At the initial stages of the data collection procedure it became evident that rudimentary data managing pipelines could have led to irrecoverable data loss. Hence, an automated data collection pipeline was developed, ensuring

constant data collection into a safe container located in a secure virtual machine, allowing for reliable operation with minimal intervention. Likewise, the processing of very large amounts of text required the use and development of computationally efficient processing pipelines that could exploit the capabilities of University of Leeds High Performance Computer (ARC4) in which most of the data were hosted. After the description of the data collection and data pre-processing steps, the chapter continues with a study of the most common emotions detection techniques and algorithms that were tested in this thesis. This section provides a description of the algorithms as well as an explanation of their mathematical foundations.

The main results are presented in Chapter 5. The chapter starts by presenting the result of the Methodological Approach Selection study, in which several techniques and algorithms for emotion detection were evaluated in terms of their predictive accuracy. Although the final results of this study favoured modern deep-learning based classification techniques, several simpler machine learning techniques performed admirably well, and depending on the context and needs of the researcher, they should be considered as viable competitive options for simple text classification tasks. The chapter continues by introducing the results of the Data Augmentation by back-translation study, conceived as a solution for the low quality of the Spanish emotion detection training dataset. This rather new technique proved to be effective in increasing the accuracy of models trained on deficient datasets. Moreover, this study also provided novel findings on the effects of using distant vs near languages in the process of data augmentation, showing that back-translation using distant languages results in loss of semantic meaning and ultimately in loss of accuracy. After this section, the chapter outlines the results of the Aspect Keyword Selection study, in which knowledge discovery techniques were used to determine which keywords represented best the four aspects in each of the three social movements studied. Finally, the chapter delves into the results of the Emotions in Social Movements study, in which the emotional trends detected in the tweets data are analysed in conjunction with the trends of reported violent and non-violent collective actions for each movement. The analysis revealed that although anger is still the main emotion in the communications of social movements, it is not an immediate predictor of collective actions, while other emotions seem to predict violent and non-violent actions. Moreover, the specific combination of target-emotion which best predicted collective actions varied widely between social movements. However, a consistent feedback loop in which reports of non-violent collective actions inhibited the occurrence of further collective actions, and reports of violent collective actions boosted the occurrence of violent collective

actions in the future was found in the data. These unexpected results highlight the necessity of further research in the field in order to fully understand the dynamics of emotions of social movements and how they influence collective action within their respective contexts.

Finally, I believe that my results and methodology provide a viable alternative to analyse social movements in an era in which their structure is becoming more diffuse and decentralised making traditional data collection methods more difficult. By using digital text data, and a simple model of the influence of emotion over violent and non-violent collective actions (see 2.5) I was able to parameterise certain key factors of social movements and I was able to measure them and analyse them robustly across very different contexts.

6.2 Final reflections

The project carried out in this thesis was intrinsically interdisciplinary. It tried to combined the knowledge of political and social psychology on social movements and collective actions and the latest developments in natural language processing to gain insights from an unstructured source of data (e.g., social media posts).

In particular, by studying in detail machine learning and deep learning algorithms I was able to understand their great potential for social science, but also what their limitations and dangers are. Machine learning, and in particular deep learning in recent years, has demonstrated to be able of great precision in relatively complex tasks, like image classification and language understanding. In particular, image classification models have become precise enough to be used by social scientist in demographics inference (gender, age and ethnicity inference) (Shin et al. 2017) and language models have became complex enough to be used in misinformation detection, hate speech detection and other classification tasks related to political text (Heidari et al. 2021). However, quality of training data has become one of the major issues in current deep learning research. Training data are essential for any kind of deep learning task, as it is the source of information for the model. Bias in the training data means biased results in the models, even if the models have excellent accuracy. What is more concerning is that complex deep learning models tend to be taken as expert systems, cementing the biases that come with the training data used to build them. These biases have been demonstrated to have a disproportionate impact on ethnic and gender minorities (Bender et al. 2021; Gebru 2020). As social scientists, I believe we are better equipped than other disciplines and areas

of study to detect and tackle these biases. Hence, I also believe we need to engage in the discussions about algorithmic biases and call out these issues whenever they are found. To do so, a basic understanding of how these algorithms and machine learning techniques operate is necessary, however, I believe the effort put in understanding them is worthy if this helps to prevent algorithmic biases and discrimination issues that will be more and more common as AI becomes more prevalent in our daily lives.

References

Abrigo, Michael RM and Love, Inessa (2016). "Estimation of panel vector autoregression in Stata". In: *The Stata Journal* 16.3, pp. 778–804.

Agarwal, Apoorv, Xie, Boyi, Vovsha, Ilia, Rambow, Owen, and Passonneau, Rebecca (June 23, 2011). "Sentiment analysis of Twitter data". In: *Proceedings of the Workshop on Languages in Social Media*. LSM '11. Portland, Oregon: Association for Computational Linguistics, pp. 30–38. ISBN: 9781932432961. (Visited on 12/02/2021).

Arnold, Magda B (1960). "Emotion and personality." In.

Arzheimer, Kai, Evans, Jocelyn, and Lewis-Beck, Michael (2017). *The SAGE Handbook of Electoral Behaviour*. 2 vols. 55 City Road, London. DOI: 10.4135/9781473957978. URL: https://sk.sagepub.com/Reference/the-sage-handbook-of-electoral-behaviour (visited on 04/18/2022).

Bakshi, Rushlene Kaur, Kaur, Navneet, Kaur, Ravneet, and Kaur, Gurpreet (Mar. 2016). "Opinion mining and sentiment analysis". In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 452–455.

Bargh, John A., Gollwitzer, Peter M., and Oettingen, Gabriele (June 30, 2010). "Motivation". In: *Handbook of Social Psychology*. Ed. by Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey. Hoboken, NJ, USA: John Wiley & Sons, Inc., socpsy001008. ISBN: 9780470561119. DOI: 10.1002/9780470561119.socpsy001008. URL: https://onlinelibrary.wiley.com/doi/10.1002/9780470561119.socpsy001008 (visited on 04/18/2022).

Bartholomew, Amy and Mayer, Margit (1992). "Nomads of the Present: Melucci's Contribution toNew Social Movement'Theory". In: *Theory, Culture & Society* 9.4, pp. 141–159.

Bayes, Thomas (1763). "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S". In: *Philosophical transactions of the Royal Society of London* 53, pp. 370–418.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (Apr. 2005). "The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers". In: *Trends in Cognitive Sciences* 9.4, 159–162, discussion 162–164. ISSN: 1364-6613. DOI: 10.1016/j.tics.2005.02.002.

Bechara, Antoine (2003). "Risky business: emotion, decision-making, and addiction". In: *Journal of Gambling Studies* 19.1, pp. 23–51. ISSN: 1050-5350. DOI: 10.1023/a:1021223113233.

Bechara, Antoine and Damasio, Antonio R. (Aug. 1, 2005). "The somatic marker hypothesis: A neural theory of economic decision". In: *Games and Economic Behavior*. Special Issue on Neuroeconomics 52.2, pp. 336–372. ISSN: 0899-8256. DOI: 10.1016/j.geb.2004.06.010. URL: https://www.sciencedirect.com/science/article/pii/S0899825604001034 (visited on 04/18/2022).

Becker, Julia C, Tausch, Nicole, and Wagner, Ulrich (2011a). "Emotional consequences of collective action participation: Differentiating self-directed and outgroup-directed emotions". In: *Personality and Social Psychology Bulletin* 37.12, pp. 1587–1598.

Becker, Julia C. and Tausch, Nicole (Jan. 1, 2015). "A dynamic model of engagement in normative and non-normative collective action: Psychological antecedents, consequences, and barriers". In: *European Review of Social Psychology* 26.1, pp. 43–92. ISSN: 1046-3283. DOI: 10.1080/10463283.2015.1094265. URL: https://doi.org/10.1080/10463283.2015.1094265 (visited on 04/18/2022).

Becker, Julia C., Tausch, Nicole, and Wagner, Ulrich (Dec. 2011b). "Emotional Consequences of Collective Action Participation: Differentiating Self-Directed and Outgroup-Directed Emotions". In: *Personality and Social Psychology Bulletin* 37.12, pp. 1587–1598. ISSN: 0146-1672,

1552-7433. DOI: 10.1177/0146167211414145. URL: http://journals.sagepub.com/doi/10.1177/0146167211414145 (visited on 04/18/2022).

Beddiar, Djamila Romaissa, Jahan, Md Saroar, and Oussalah, Mourad (2021). "Data expansion using back translation and paraphrasing for hate speech detection". In: *Online Social Networks and Media* 24, p. 100153.

Bender, Emily M, Gebru, Timnit, McMillan-Major, Angelina, and Shmitchell, Shmargaret (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623.

Benford, Robert D. and Snow, David A. (Aug. 2000). "Framing Processes and Social Movements: An Overview and Assessment". In: *Annual Review of Sociology* 26.1, pp. 611–639. ISSN: 0360-0572, 1545-2115. DOI: 10.1146/annurev.soc.26.1.611. URL: https://www.annualreviews.org/doi/10.1146/annurev.soc.26.1.611 (visited on 04/18/2022).

Bennett, W. (Jan. 1, 2003). "Communicating Global Activism". In: *Information, Communication & Society* 6.2, pp. 143–168. ISSN: 1369-118X. DOI: 10.1080/1369118032000093860a. URL: https://doi.org/10.1080/1369118032000093860a (visited on 04/18/2022).

Berridge, Kent C (Apr. 1, 2004). "Motivation concepts in behavioral neuroscience". In: *Physiology & Behavior*. Reviews on Ingestive Science 81.2, pp. 179–209. ISSN: 0031-9384. DOI: 10.1016/j.physbeh.2004.02.004. URL: https://www.sciencedirect.com/science/article/pii/S0031938404000435 (visited on 04/18/2022).

Binali, Haji, Wu, Chen, and Potdar, Vidyasagar (2010). "Computational approaches for emotion detection in text". In: *4th IEEE international conference on digital ecosystems and technologies*. IEEE, pp. 172–177.

Blei, David M, Ng, Andrew Y, and Jordan, Michael I (2003). "Latent dirichlet allocation". In: *the Journal of machine Learning research* 3, pp. 993–1022.

Bonenfant, Maude and Meurs, Marie-Jean (2020). "Collaboration Between Social Sciences and Computer Science: Toward a Cross-Disciplinary Methodology for Studying Big Social Data from Online Communities". In: *Second International Handbook of Internet Research*. Ed. by

Jeremy Hunsinger, Matthew M. Allen, and Lisbeth Klastrup. Dordrecht: Springer Netherlands, pp. 47–63. ISBN: 9789402415551. DOI: 10.1007/978-94-024-1555-1_39. URL: https://doi.org/10.1007/978-94-024-1555-1_39 (visited on 11/24/2021).

Boser, Bernhard E., Guyon, Isabelle M., and Vapnik, Vladimir N. (July 1, 1992). "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory*. COLT '92. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, pp. 144–152. ISBN: 9780897914970. DOI: 10.1145/130385.130401. URL: https://doi.org/10.1145/130385.130401 (visited on 12/05/2021).

Bosse, Tibor, Jonker, Catholijn M., and Treur, Jan (Mar. 1, 2008). "Formalisation of Damasio's theory of emotion, feeling and core consciousness". In: *Consciousness and Cognition* 17.1, pp. 94–113. ISSN: 1053-8100. DOI: 10.1016/j.concog.2007.06.006. URL: https://www.sciencedirect.com/science/article/pii/S1053810007000633 (visited on 04/18/2022).

Bridle, John S (1990). "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition". In: *Neurocomputing*. Springer, pp. 227–236.

Bruns, Stephan B and Stern, David I (2019). "Lag length selection and p-hacking in Granger causality testing: prevalence and performance of meta-regression models". In: *Empirical Economics* 56.3, pp. 797–830.

Brym, Robert, Slavina, Anna, Todosijevic, Mina, and Cowan, David (Nov. 2018). "Social Movement Horizontality in the Internet Age? A Critique of Castells in Light of the Trump Victory". In: *Canadian Review of Sociology = Revue Canadienne De Sociologie* 55.4, pp. 624–634. ISSN: 1755-618X. DOI: 10.1111/cars.12219.

Burkholder, Gary J. and Harlow, Lisa L. (July 1, 2003). "An Illustration of a Longitudinal Cross-Lagged Design for Larger Structural Equation Models". In: *Structural Equation Modeling: A Multidisciplinary Journal* 10.3, pp. 465–486. ISSN: 1070-5511. DOI: 10.1207/S15328007SEM1003_8. URL: https://doi.org/10.1207/S15328007SEM1003_8 (visited on 04/18/2022).

Castells, Manuel (Jan. 26, 2010). *End of Millennium*. Google-Books-ID: 1wDLJAGDRGYC. John Wiley & Sons. 489 pp. ISBN: 9781444323443.

– (2015). *Networks of outrage and hope: Social movements in the Internet age*. John Wiley & Sons.

Cheng, Kris (July 21, 2016). *Hong Kong Occupy activists including Joshua Wong guilty of unlawful assembly*. Hong Kong Free Press HKFP. URL: https://hongkongfp.com/2016/07/21/breaking-hong-kong-occupy-activists-including-joshua-wong-guilty-of-unlawful-assembly/ (visited on 04/28/2022).

Chernin, Kelly (June 17, 2019). *Mass protests protect Hong Kong's legal autonomy from China – for now*. The Conversation. URL: http://theconversation.com/mass-protests-protect-hong-kongs-legal-autonomy-from-china-for-now-118753 (visited on 04/28/2022).

Cheung, Helier (June 17, 2019). "Hong Kong extradition: How radical youth forced the government's hand". In: *BBC News*. URL: https://www.bbc.com/news/world-asia-china-48655474 (visited on 04/28/2022).

Cho, Kyunghyun, Van Merrienboer, Bart, Gulcehre, Caglar, Bahdanau, Dzmitry, Bougares, Fethi, Schwenk, Holger, and Bengio, Yoshua (Sept. 2, 2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". In: *arXiv:1406.1078 [cs, stat]*. arXiv: 1406.1078. URL: http://arxiv.org/abs/1406.1078 (visited on 12/07/2021).

Choma, Becky, Hodson, Gordon, Jagayat, Arvin, and Hoffarth, Mark R (2020). "Right-wing ideology as a predictor of collective action: A test across four political issue domains". In: *Political Psychology* 41.2, pp. 303–322.

Coase, Ronald H. (1977). "Economics and Contiguous Disciplines". In: *The Organization and Retrieval of Economic Knowledge: Proceedings of a Conference held by the International Economic Association at Kiel, West Germany*. Ed. by Mark Perlman. London: Palgrave Macmillan UK, pp. 481–495. ISBN: 9781349033256. DOI: 10.1007/978-1-349-03325-6_26. URL: https://doi.org/10.1007/978-1-349-03325-6_26 (visited on 04/18/2022).

Cryer, Jonathan D (1986). *Time series analysis*. Vol. 286. Springer.

Csikszentmihalyi, Mihaly and Seligman, M (2000). "Positive psychology". In: *American psychologist* 55.1, pp. 5–14.

Dalgleish, Tim (July 2004). "The emotional brain". In: *Nature Reviews Neuroscience* 5.7, pp. 583–589. ISSN: 1471-0048. DOI: 10.1038/nrn1432. URL: https://www.nature.com/articles/nrn1432 (visited on 04/18/2022).

Damasio, A. (1998). "The somatic marker hypothesis and the possible functions of the prefrontal cortex". In: *The Prefrontal Cortex*. Oxford: Oxford University Press. ISBN: 9780198524410. DOI: 10.1093/acprof:oso/9780198524410.003.0004. URL: https://oxford.universitypressscholarshi com/10.1093/acprof:oso/9780198524410.001.0001/acprof-9780198524410-chapter-4 (visited on 04/18/2022).

– (Oct. 5, 2000). *The Feeling Of What Happens: Body, Emotion and the Making of Consciousness*. Reprint edition. London: Vintage. 400 pp. ISBN: 9780099288763.

Damasio, A., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., and Hichwa, R. D. (Oct. 2000). "Subcortical and cortical brain activity during the feeling of self-generated emotions". In: *Nature Neuroscience* 3.10, pp. 1049–1056. ISSN: 1097-6256. DOI: 10.1038/79871.

De Moor, Joost, Uba, Katrin, Wahlström, Mattias, Wennerhag, Magnus, and De Vydt, Michiel (2020). *Protest for a future II: Composition, mobilization and motives of the participants in Fridays For Future climate protests on 20-27 September, 2019, in 19 cities around the world.*

De Rivera, Joseph (1992). "Emotional climate: Social structure and emotional dynamics." In: *A preliminary draft of this chapter was discussed at a workshop on emotional climate sponsored by the Clark European Center in Luxembourg on Jul 12–14, 1991.* John Wiley & Sons.

Deci, Edward L. and Ryan, Richard M. (Jan. 30, 2010). "Intrinsic Motivation". In: *The Corsini Encyclopedia of Psychology*. Ed. by Irving B. Weiner and W. Edward Craighead. Hoboken, NJ, USA: John Wiley & Sons, Inc., corpsy0467. ISBN: 9780470479216. DOI: 10.1002/9780470479216.corpsy0467. URL: https://onlinelibrary.wiley.com/doi/10.1002/9780470479216.corpsy0467 (visited on 04/18/2022).

Delgado, M. R., Olsson, A., and Phelps, E. A. (July 2006). "Extending animal models of fear conditioning to humans". In: *Biological Psychology* 73.1, pp. 39–48. ISSN: 0301-0511. DOI: 10.1016/j.biopsycho.2006.01.006.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, and Toutanova, Kristina (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.

Dickey, David A and Fuller, Wayne A (1979). "Distribution of the estimators for autoregressive time series with a unit root". In: *Journal of the American statistical association* 74.366a, pp. 427–431.

Djatmiko, Fahim, Ferdiana, Ridi, and Faris, Muhammad (Mar. 2019). "A Review of Sentiment Analysis for Non-English Language". In: *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*. 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), pp. 448–451. DOI: 10.1109/ICAIIT.2019.8834552.

Drury, J. and Reicher, S. (Oct. 1999). "The Intergroup Dynamics of Collective Empowerment: Substantiating the Social Identity Model of Crowd Behavior". In: *Group Processes & Intergroup Relations* 2.4, pp. 381–402. ISSN: 1368-4302, 1461-7188. DOI: 10.1177/1368430299024005. URL: http://journals.sagepub.com/doi/10.1177/1368430299024005 (visited on 04/18/2022).

– (Jan. 2005). "Explaining enduring empowerment: a comparative study of collective action and psychological outcomes". In: *European Journal of Social Psychology* 35.1, pp. 35–58. ISSN: 0046-2772, 1099-0992. DOI: 10.1002/ejsp.231. URL: https://onlinelibrary.wiley.com/doi/10.1002/ejsp.231 (visited on 04/18/2022).

– (Dec. 2009). "Collective Psychological Empowerment as a Model of Social Change: Researching Crowds and Power". In: *Journal of Social Issues* 65.4, pp. 707–725. ISSN: 00224537, 15404560. DOI: 10.1111/j.1540-4560.2009.01622.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1540-4560.2009.01622.x (visited on 04/18/2022).

DW (Aug. 20, 2019). *Twitter accuses China of anti-Hong Kong protest campaign*. dw.com. URL: https://www.dw.com/en/twitter-accuses-china-of-anti-hong-kong-protest-campaign/a-50089240 (visited on 11/14/2022).

Edelmann, Achim, Wolff, Tom, Montagne, Danielle, and Bail, Christopher A (2020). "Computational social science and sociology". In: *Annual Review of Sociology* 46.1, p. 61.

Egan, S. K. and Perry, D. G. (July 2001). "Gender identity: a multidimensional analysis with implications for psychosocial adjustment". In: *Developmental Psychology* 37.4, pp. 451–463. ISSN: 0012-1649. DOI: 10.1037//0012-1649.37.4.451.

Ekman, P. (1992). "An argument for basic emotions". In: *Cognition & emotion* 6.3-4, pp. 169–200.

– (1999). "Basic emotions". In: *Handbook of cognition and emotion* 98.45-60, p. 16.

Ekman, P. and Friesen, W. V. (Feb. 1971). "Constants across cultures in the face and emotion". In: *Journal of Personality and Social Psychology* 17.2, pp. 124–129. ISSN: 0022-3514. DOI: 10.1037/h0030377.

Ellsworth, P. C. (Apr. 1994). "William James and emotion: is a century of fame worth a century of misunderstanding?" In: *Psychological Review* 101.2, pp. 222–229. ISSN: 0033-295X. DOI: 10.1037/0033-295x.101.2.222.

Erikson, Erik H. (1993). *Childhood and society*. New York: Norton. 445 pp. ISBN: 9780393310689.

Fanselow, Michael S. (Sept. 1, 1990). "Factors governing one-trial contextual conditioning". In: *Animal Learning & Behavior* 18.3, pp. 264–270. ISSN: 1532-5830. DOI: 10.3758/BF03205285. URL: https://doi.org/10.3758/BF03205285 (visited on 04/18/2022).

Feddes, Allard R., Mann, Liesbeth, and Doosje, Bertjan (Dec. 2012). "From extreme emotions to extreme actions: explaining non-normative collective action and reconciliation". In: *The Behavioral and Brain Sciences* 35.6, pp. 432–433. ISSN: 1469-1825. DOI: 10.1017/S0140525X12001197.

Fernandez, Adriela and Vera, Marisol (July 2012). "The Bachelet Presidency and the End of Chile's Concertación Era". In: *Latin American Perspectives* 39.4, pp. 5–18. ISSN: 0094-582X,

1552-678X. DOI: 10.1177/0094582X12442054. URL: http://journals.sagepub.com/doi/10.1177/0094582X12442054 (visited on 05/12/2022).

FFF (2022). *Fridays For Future – Our demands. Act now!* Fridays For Future. URL: https://fridaysforfuture.org/what-we-do/our-demands/ (visited on 05/11/2022).

Flor, Herta and Birbaumer, Niels (Jan. 1, 2015). "Fear Conditioning: Overview". In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Ed. by James D. Wright. Oxford: Elsevier, pp. 849–853. ISBN: 9780080970875. DOI: 10.1016/B978-0-08-097086-8.55022-3. URL: https://www.sciencedirect.com/science/article/pii/B9780080970868550223 (visited on 04/18/2022).

Ford, Martin (1992). *Motivating Humans: Goals, Emotions, and Personal Agency Beliefs*. Thousand Oaks, California. DOI: 10.4135/9781483325361. URL: https://sk.sagepub.com/books/motivating-humans (visited on 04/18/2022).

Fourcade, Marion, Ollion, Etienne, and Algan, Yann (Feb. 2015). "The Superiority of Economists". In: *Journal of Economic Perspectives* 29.1, pp. 89–114. ISSN: 0895-3309. DOI: 10.1257/jep.29.1.89. URL: https://www.aeaweb.org/articles?id=10.1257/jep.29.1.89 (visited on 04/18/2022).

Freeman, John R, Williams, John T, and Lin, Tse-min (1989). "Vector autoregression and the study of politics". In: *American Journal of Political Science*, pp. 842–877.

Freeman, Richard B. (Sept. 1999). "It's Better Being an Economist (But Don't Tell Anyone)." In: *Journal of Economic Perspectives* 13.3, pp. 139–145. ISSN: 0895-3309. DOI: 10.1257/jep.13.3.139. URL: https://www.aeaweb.org/articles?id=10.1257/jep.13.3.139 (visited on 04/18/2022).

Gatti, Lorenzo, Guerini, Marco, and Turchi, Marco (Oct. 2016). "SentiWords: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis". In: *IEEE Transactions on Affective Computing* 7.4, pp. 409–421. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2015.2476456.

Gebru, Timnit (2020). "Race and gender". In: *The Oxford handbook of ethics of aI*, pp. 251–269.

Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron (Nov. 18, 2016). "Deep Feedforward Networks". In: *Deep Learning*. Red. by Francis Bach. Adaptive Computation and Machine Learning series. Cambridge, MA, USA: MIT Press, pp. 163–220. ISBN: 9780262035613.

Goodwin, Jeff and Jasper, James M (1999). "Caught in a winding, snarling vine: The structural bias of political process theory". In: *Sociological forum*. Vol. 14. 1. Springer, pp. 27–54.

Google (2022). *Imbalanced Data | Data Preparation and Feature Engineering for Machine Learning*. Google Developers. URL: https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data (visited on 03/12/2022).

Granger, Clive WJ (1969). "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: journal of the Econometric Society*, pp. 424–438.

Grinin, Leonid, Korotayev, Andrey, and Tausch, Arno (2019). "Introduction. Why Arab Spring Became Arab Winter". In: *Islamism, Arab Spring, and the Future of Democracy: World System and World Values Perspectives*. Ed. by Leonid Grinin, Andrey Korotayev, and Arno Tausch. Cham: Springer International Publishing, pp. 1–24. ISBN: 9783319910772. DOI: 10.1007/978-3-319-91077-2_1. URL: https://doi.org/10.1007/978-3-319-91077-2_1 (visited on 04/18/2022).

Gu, Simeng, Wang, Fushun, Patel, Nitesh P, Bourgeois, James A, and Huang, Jason H (2019). "A model for basic emotions using observations of behavior in Drosophila". In: *Frontiers in psychology*, p. 781.

Gurr, Ted Robert (Nov. 16, 2015). *Why Men Rebel*. New York: Routledge. 440 pp. ISBN: 9781315631073. DOI: 10.4324/9781315631073.

Hareli, Shlomo, Kafetsios, Konstantinos, and Hess, Ursula (2015). "A cross-cultural study on emotion expression and the learning of social norms". In: *Frontiers in psychology* 6, p. 1501.

He, Haibo and Garcia, Edwardo A (2009). "Learning from imbalanced data". In: *IEEE Transactions on knowledge and data engineering* 21.9, pp. 1263–1284.

Head, Megan L, Holman, Luke, Lanfear, Rob, Kahn, Andrew T, and Jennions, Michael D (2015). "The extent and consequences of p-hacking in science". In: *PLoS biology* 13.3, e1002106.

Heidari, Maryam, Zad, Samira, Hajibabaee, Parisa, Malekzadeh, Masoud, HekmatiAthar, Seyyed-Pooya, Uzuner, Ozlem, and Jones, James H (2021). "Bert model for fake news detection based on social bot activities in the covid-19 pandemic". In: *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, pp. 0103–0109.

Heyes, Cecilia (2019). "Précis of cognitive gadgets: The cultural evolution of thinking". In: *Behavioral and Brain Sciences* 42.

Hochreiter, Sepp and Schmidhuber, Jürgen (1996). "LSTM can solve hard long time lag problems". In: *Advances in neural information processing systems* 9, pp. 473–479.

Hofman, Jake M, Watts, Duncan J, Athey, Susan, Garip, Filiz, Griffiths, Thomas L, Kleinberg, Jon, Margetts, Helen, Mullainathan, Sendhil, Salganik, Matthew J, Vazire, Simine, et al. (2021). "Integrating explanation and prediction in computational social science". In: *Nature* 595.7866, pp. 181–188.

Holbig, Heike (2020). "Be water, my friend: Hong Kong's 2019 Anti-extradition protests". In: *International Journal of Sociology* 50.4, pp. 325–337.

Hornik, Kurt, Stinchcombe, Maxwell, and White, Halbert (1989). "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5, pp. 359–366.

Hughes, Brent L. and Zaki, Jamil (Feb. 1, 2015). "The neuroscience of motivated cognition". In: *Trends in Cognitive Sciences* 19.2, pp. 62–64. ISSN: 1364-6613, 1879-307X. DOI: 10.1016/j.tics.2014.12.006. URL: https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(14)00270-8 (visited on 04/18/2022).

IPCC (2014). *AR5 Climate Change 2014: Impacts, Adaptation, and Vulnerability — IPCC*. URL: https://www.ipcc.ch/report/ar5/wg2/ (visited on 01/01/2014).

– (2022). *AR6 Climate Change 2022: Impacts, Adaptation and Vulnerability — IPCC*. URL: https://www.ipcc.ch/report/sixth-assessment-report-working-group-ii/ (visited on 01/01/2022).

Isaak, Jim and Hanna, Mina J. (Aug. 2018). "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection". In: *Computer* 51.8, pp. 56–59. ISSN: 1558-0814. DOI: 10.1109/MC.2018.3191268.

Ivanov, Ventzislav and Kilian, Lutz (2005). "A practitioner's guide to lag order selection for VAR impulse response analysis". In: *Studies in Nonlinear Dynamics & Econometrics* 9.1.

Jack, Rachael E, Garrod, Oliver GB, and Schyns, Philippe G (2014). "Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time". In: *Current biology* 24.2, pp. 187–192.

Jack, Rachael E, Garrod, Oliver GB, Yu, Hui, Caldara, Roberto, and Schyns, Philippe G (2012). "Facial expressions of emotion are not culturally universal". In: *Proceedings of the National Academy of Sciences* 109.19, pp. 7241–7244.

Jack, Rachael E, Sun, Wei, Delis, Ioannis, Garrod, Oliver GB, and Schyns, Philippe G (2016). "Four not six: Revealing culturally common facial expressions of emotion." In: *Journal of Experimental Psychology: General* 145.6, p. 708.

James, William (1884). "II.—WHAT IS AN EMOTION ?" In: *Mind* os-IX.34, pp. 188–205. ISSN: 0026-4423, 1460-2113. DOI: 10.1093/mind/os-IX.34.188. URL: https://academic.oup.com/mind/article-lookup/doi/10.1093/mind/os-IX.34.188 (visited on 04/18/2022).

James, William, Burkhardt, Frederick, Bowers, Fredson, and Skrupskelis, Ignas K (1890). *The principles of psychology*. Vol. 1. 2. Macmillan London.

Jenkins, J Craig (1983). "Resource mobilization theory and the study of social movements". In: *Annual review of sociology*, pp. 527–553.

Jordan, M. I. and Mitchell, T. M. (July 17, 2015). "Machine learning: Trends, perspectives, and prospects". In: *Science*. DOI: 10.1126/science.aaa8415. URL: https://www.science.org/doi/abs/10.1126/science.aaa8415 (visited on 12/03/2021).

Kadhim, Ammar Ismael (June 1, 2019). "Survey on supervised machine learning techniques for automatic text classification". In: *Artificial Intelligence Review* 52.1, pp. 273–292. ISSN:

1573-7462. DOI: 10.1007/s10462-018-09677-1. URL: https://doi.org/10.1007/s10462-018-09677-1 (visited on 12/05/2021).

Kaity, Mohammed and Balakrishnan, Vimala (Dec. 1, 2020). "Sentiment lexicons and non-English languages: a survey". In: *Knowledge and Information Systems* 62.12, pp. 4445–4480. ISSN: 0219-3116. DOI: 10.1007/s10115-020-01497-6. URL: https://doi.org/10.1007/s10115-020-01497-6 (visited on 05/21/2022).

Karami, Amir, Lundy, Morgan, Webb, Frank, and Dwivedi, Yogesh K. (2020). "Twitter and Research: A Systematic Literature Review Through Text Mining". In: *IEEE Access* 8, pp. 67698–67717. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2983656.

Kemp, Simon (2021). *Digital 2021 April Statshot Report*. DataReportal – Global Digital Insights. URL: https://datareportal.com/reports/digital-2021-april-global-statshot (visited on 11/25/2021).

Kenny, David A. (Sept. 29, 2014). "Cross-Lagged Panel Design". In: *Wiley StatsRef: Statistics Reference Online*. Ed. by N. Balakrishnan, Theodore Colton, Brian Everitt, Walter Piegorsch, Fabrizio Ruggeri, and Jozef L. Teugels. 1st ed. Wiley. ISBN: 9781118445112. DOI: 10.1002/9781118445112.stat06464. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06464 (visited on 04/18/2022).

Kerr, Norbert L (1998). "HARKing: Hypothesizing after the results are known". In: *Personality and social psychology review* 2.3, pp. 196–217.

Khondker, Habibul Haque (Oct. 1, 2011). "Role of the New Media in the Arab Spring". In: *Globalizations* 8.5, pp. 675–679. ISSN: 1474-7731. DOI: 10.1080/14747731.2011.621287. URL: https://doi.org/10.1080/14747731.2011.621287 (visited on 04/18/2022).

Khoo, Christopher SG and Johnkhan, Sathik Basha (Aug. 1, 2018). "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons". In: *Journal of Information Science* 44.4, pp. 491–511. ISSN: 0165-5515. DOI: 10.1177/0165551517703514. URL: https://doi.org/10.1177/0165551517703514 (visited on 12/03/2021).

Klandermans, Bert and Stekelenburg, Jacquelien Van (Sept. 4, 2013). *Social Movements and the Dynamics of Collective Action*. The Oxford Handbook of Political Psychology. DOI: 10.

1093/oxfordhb/9780199760107.013.0024. URL: https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199760107.001.0001/oxfordhb-9780199760107-e-024 (visited on 04/18/2022).

Kobayashi, Tetsuro, Song, Jaehyun, and Chan, Polly (2021). "Does repression undermine opposition demands? The case of the Hong Kong National Security Law". In: *Japanese Journal of Political Science* 22.4, pp. 268–286.

Kouloumpis, Efthymios, Wilson, Theresa, and Moore, Johanna (2011). "Twitter Sentiment Analysis: The Good the Bad and the OMG!" In: *Proceedings of the International AAAI Conference on Web and Social Media* 5.1, pp. 538–541. ISSN: 2334-0770. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14185 (visited on 12/02/2021).

Kozak, Piotr (Jan. 23, 2018). "Chile president-elect reveals hardline cabinet with ties to Pinochet". In: *The Guardian*. ISSN: 0261-3077. URL: https://www.theguardian.com/world/2018/jan/23/chile-president-elect-sebastian-pinera-andres-chadwick (visited on 05/12/2022).

Kuo, Lily (Aug. 18, 2019). "Hong Kong's dilemma: fight or resist peacefully". In: *The Observer*. ISSN: 0029-7712. URL: https://www.theguardian.com/world/2019/aug/18/hong-kong-protesters-dilemma-fight-resist-peacefully-china-troops (visited on 05/01/2022).

Lague, David, Pomfret, James, and Torode, Greg (Dec. 20, 2019). "How murder, kidnappings and a miscalculation set off Hong Kong revolt". In: *Reuters*. URL: https://www.reuters.com/investigates/special-report/hongkong-protests-extradition-narrative/ (visited on 04/27/2022).

Lam, Jeffie, Lok-kei, Sum, and Leung, Kanis (Oct. 3, 2019). *Was police officer justified in opening fire on Hong Kong protester?* South China Morning Post. URL: https://www.scmp.com/news/hong-kong/politics/article/3031325/hong-kong-protests-was-police-officer-justified-opening (visited on 05/01/2022).

Landmann, Helen and Rohmann, Anette (2020). "Being moved by protest: Collective efficacy beliefs and injustice appraisals enhance collective action intentions for forest protection via positive and negative emotions". In: *Journal of Environmental Psychology* 71, p. 101491.

Lantos, Nóra Anna, Kende, Anna, Becker, Julia C, and McGarty, Craig (2020). "Pity for economically disadvantaged groups motivates donation and ally collective action intentions". In: *European Journal of Social Psychology* 50.7, pp. 1478–1499.

Lazer, David, Pentland, Alex, Adamic, Lada, Aral, Sinan, Barabási, Albert-László, Brewer, Devon, Christakis, Nicholas, Contractor, Noshir, Fowler, James, Gutmann, Myron, et al. (2009). "Computational social science". In: *Science* 323.5915, pp. 721–723.

Leary, Mark R. (2004). *The Curse of the Self: Self-Awareness, Egotism, and the Quality of Human Life*. New York: Oxford University Press. 237 pp. ISBN: 9780195172423. DOI: 10.1093/acprof:oso/9780195172423.001.0001. URL: https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780195172423.001.0001/acprof-9780195172423 (visited on 04/18/2022).

Leary, Mark R. and Tangney, June Price (Dec. 21, 2011). *Handbook of Self and Identity, Second Edition*. Google-Books-ID: VukSQuVMQy0C. Guilford Press. 769 pp. ISBN: 9781462503124.

LeBlanc, Vicki R., McConnell, Meghan M., and Monteiro, Sandra D. (Mar. 1, 2015). "Predictable chaos: a review of the effects of emotions on attention, memory and decision making". In: *Advances in Health Sciences Education* 20.1, pp. 265–282. ISSN: 1573-1677. DOI: 10.1007/s10459-014-9516-6. URL: https://doi.org/10.1007/s10459-014-9516-6 (visited on 04/18/2022).

LeCun, Yann, Bengio, Yoshua, and Hinton, Geoffrey (May 2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. URL: https://www.nature.com/articles/nature14539 (visited on 12/07/2021).

Leung, Hillary (Aug. 27, 2019). *Then and Now: 79 Days of Protest in Hong Kong*. Time. URL: https://time.com/5661211/hong-kong-protests-79-days/ (visited on 04/28/2022).

Levine, Linda J. and Safer, Martin A. (Oct. 2002). "Sources of Bias in Memory for Emotions". In: *Current Directions in Psychological Science* 11.5, pp. 169–173. ISSN: 0963-7214, 1467-8721. DOI: 10.1111/1467-8721.00193. URL: http://journals.sagepub.com/doi/10.1111/1467-8721.00193 (visited on 04/18/2022).

Li, Shan and Deng, Weihong (2020). "Deep facial expression recognition: A survey". In: *IEEE transactions on affective computing*.

Liu, Pei, Wang, Xuemin, Xiang, Chao, and Meng, Weiye (2020). "A survey of text data augmentation". In: *2020 International Conference on Computer Communication and Network Security (CCNS)*. IEEE, pp. 191–195.

Liu, Yinhan, Ott, Myle, Goyal, Naman, Du, Jingfei, Joshi, Mandar, Chen, Danqi, Levy, Omer, Lewis, Mike, Zettlemoyer, Luke, and Stoyanov, Veselin (2019). "Roberta: A robustly optimized bert pretraining approach". In: *arXiv preprint arXiv:1907.11692*.

Lopez, Luciano and Weber, Sylvain (2017). "Testing for Granger causality in panel data". In: *The Stata Journal* 17.4, pp. 972–984.

Loveman, Mara (Sept. 1, 1998). "High-Risk Collective Action: Defending Human Rights in Chile, Uruguay, and Argentina". In: *American Journal of Sociology* 104.2, pp. 477–525. ISSN: 0002-9602. DOI: 10.1086/210045. URL: https://www.journals.uchicago.edu/doi/10.1086/210045 (visited on 04/18/2022).

Lütkepohl, Helmut (2013). "Vector autoregressive models". In: *Handbook of Research Methods and Applications in Empirical Macroeconomics*. Edward Elgar Publishing, pp. 139–164.

Mackie, Diane M, Smith, Eliot R, and Ray, Devin G (2008). "Intergroup emotions and intergroup relations". In: *Social and Personality Psychology Compass* 2.5, pp. 1866–1880.

Marcus, George E., Neuman, W. Russell, and MacKuen, Michael (2022). *Affective Intelligence and Political Judgment*. Chicago, IL: University of Chicago Press. 200 pp. URL: https://press.uchicago.edu/ucp/books/book/chicago/A/bo3636531.html (visited on 04/18/2022).

McAdam, Doug (July 1, 1986). "Recruitment to High-Risk Activism: The Case of Freedom Summer". In: *American Journal of Sociology* 92.1, pp. 64–90. ISSN: 0002-9602. DOI: 10.1086/228463. URL: https://www.journals.uchicago.edu/doi/10.1086/228463 (visited on 04/18/2022).

McCallum, Andrew, Nigam, Kamal, et al. (1998). "A comparison of event models for naive bayes text classification". In: *AAAI-98 workshop on learning for text categorization*. Vol. 752. 1. Citeseer, pp. 41–48.

McCarthy, John D. and Zald, Mayer N. (May 1, 1977). "Resource Mobilization and Social Movements: A Partial Theory". In: *American Journal of Sociology* 82.6, pp. 1212–1241. ISSN: 0002-9602. DOI: 10.1086/226464. URL: https://www.journals.uchicago.edu/doi/10.1086/226464 (visited on 04/18/2022).

Mellon, Jonathan and Prosser, Christopher (July 1, 2017). "Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users". In: *Research & Politics* 4.3, p. 2053168017720008. ISSN: 2053-1680. DOI: 10.1177/2053168017720008. URL: https://doi.org/10.1177/2053168017720008 (visited on 11/25/2021).

Mesquita, Batja, Boiger, Michael, and De Leersnyder, Jozefien (2016). "The cultural construction of emotions". In: *Current opinion in psychology* 8, pp. 31–36.

Metsis, Vangelis, Androutsopoulos, Ion, and Paliouras, Georgios (2006). "Spam filtering with naive bayes-which naive bayes?" In: *CEAS*. Vol. 17. Mountain View, CA, pp. 28–69.

Miller, Daniel A, Cronin, Tracey, Garcia, Amber L, and Branscombe, Nyla R (2009). "The relative impact of anger and efficacy on collective action is affected by feelings of fear". In: *Group Processes & Intergroup Relations* 12.4, pp. 445–462.

Mohammad, Saif, Bravo-Marquez, Felipe, Salameh, Mohammad, and Kiritchenko, Svetlana (June 2018). "SemEval-2018 Task 1: Affect in Tweets". In: *Proceedings of The 12th International Workshop on Semantic Evaluation*. New Orleans, Louisiana: Association for Computational Linguistics, pp. 1–17. DOI: 10.18653/v1/S18-1001. URL: https://aclanthology.org/S18-1001.

Mohammad, Saif M. (2017). "Challenges in Sentiment Analysis". In: *A Practical Guide to Sentiment Analysis*. Ed. by Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. Socio-Affective Computing. Cham: Springer International Publishing, pp. 61–83. ISBN:

9783319553948. DOI: 10.1007/978-3-319-55394-8_4. URL: https://doi.org/10.1007/978-3-319-55394-8_4 (visited on 12/03/2021).

Mohammad, Saif M. and Turney, Peter D. (June 5, 2010). "Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon". In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. CAAGET '10. Los Angeles, California: Association for Computational Linguistics, pp. 26–34. (Visited on 12/02/2021).

Murthy, Sreerama K (1998). "Automatic construction of decision trees from data: A multi-disciplinary survey". In: *Data mining and knowledge discovery* 2.4, pp. 345–389.

Nevett, Joshua (Aug. 28, 2019). "Greta Thunberg: Why are young climate activists facing so much hate?" In: *BBC News*. URL: https://www.bbc.com/news/world-49291464 (visited on 06/26/2022).

Niedźwiecki, Maciej and Ciołek, Marcin (2017). "Akaike's final prediction error criterion revisited". In: *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, pp. 237–242.

Nielsen, Finn Årup (Mar. 15, 2011). "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs". In: *arXiv:1103.2903 [cs]*. arXiv: 1103.2903. URL: http://arxiv.org/abs/1103.2903 (visited on 12/02/2021).

Olson, Mancur (1971). *The Logic of Collective Action*. Vol. 178. Harvard University Press.

Pak, Alexander and Paroubek, Patrick (May 2010). "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. LREC 2010. Valletta, Malta: European Language Resources Association (ELRA). URL: http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf (visited on 12/02/2021).

Pakpahan, Eduwin, Hoffmann, Rasmus, and Kröger, Hannes (Jan. 2, 2017). "Statistical methods for causal analysis in life course research: an illustration of a cross-lagged structural equation model, a latent growth model, and an autoregressive latent trajectories model". In: *International Journal of Social Research Methodology* 20.1, pp. 1–19. ISSN: 1364-5579. DOI:

10.1080/13645579.2015.1091641. URL: https://doi.org/10.1080/13645579.2015.1091641 (visited on 04/18/2022).

Pan, Sinno Jialin and Yang, Qiang (2009). "A survey on transfer learning". In: *IEEE Transactions on knowledge and data engineering* 22.10, pp. 1345–1359.

Parvizi, J. and Damasio, A. (Apr. 2001). "Consciousness and the brainstem". In: *Cognition* 79.1, pp. 135–160. ISSN: 0010-0277. DOI: 10.1016/s0010-0277(00)00127-x.

Parvizi, Josef, Van Hoesen, Gary W., Buckwalter, Joseph, and Damasio, Antonio (Jan. 31, 2006). "Neural connections of the posteromedial cortex in the macaque". In: *Proceedings of the National Academy of Sciences* 103.5, pp. 1563–1568. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0507729103. URL: https://pnas.org/doi/full/10.1073/pnas.0507729103 (visited on 04/18/2022).

Perkel, Jeffrey M. (Dec. 1, 2020). "Why scientists are turning to Rust". In: *Nature* 588.7836, pp. 185–186. DOI: 10.1038/d41586-020-03382-2. URL: https://www.nature.com/articles/d41586-020-03382-2 (visited on 11/25/2021).

Phillips, David P (1974). "The influence of suggestion on suicide: Substantive and theoretical implications of the Werther effect". In: *American sociological review*, pp. 340–354.

Phinney, J. S. (Nov. 1990). "Ethnic identity in adolescents and adults: review of research". In: *Psychological Bulletin* 108.3, pp. 499–514. ISSN: 0033-2909. DOI: 10.1037/0033-2909.108.3.499.

Purbrick, Martin (Aug. 8, 2019). "A Report of the 2019 Hong Kong Protests". In: *Asian Affairs* 50.4, pp. 465–487. ISSN: 0306-8374. DOI: 10.1080/03068374.2019.1672397. URL: https://doi.org/10.1080/03068374.2019.1672397 (visited on 05/01/2022).

Quoidbach, Jordi, Gilbert, Daniel T, and Wilson, Timothy D (2013). "The end of history illusion". In: *science* 339.6115, pp. 96–98.

Reddy, Martin (Jan. 1, 2011). "Chapter 1 - Introduction". In: *API Design for C++*. Ed. by Martin Reddy. Boston: Morgan Kaufmann, pp. 1–19. ISBN: 9780123850034. DOI: 10.1016/

B978-0-12-385003-4.00001-4. URL: https://www.sciencedirect.com/science/article/pii/B9780123850034000014 (visited on 11/26/2021).

Reeve, Johnmarshall (2018). *Understanding motivation and emotion*. John Wiley & Sons.

Reicher, S. (Dec. 1996). "'The Crowd' century: Reconciling practical success with theoretical failure". In: *British Journal of Social Psychology* 35.4, pp. 535–553. ISSN: 01446665. DOI: 10.1111/j.2044-8309.1996.tb01113.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.2044-8309.1996.tb01113.x (visited on 04/18/2022).

Rennie, Jason D, Shih, Lawrence, Teevan, Jaime, and Karger, David R (2003). "Tackling the poor assumptions of naive bayes text classifiers". In: *Proceedings of the 20th international conference on machine learning (ICML-03)*, pp. 616–623.

Reporters, SCMP (June 16, 2019). *As it happened: Hong Kong's historic march as '2 million' people peacefully protest*. South China Morning Post. URL: https://www.scmp.com/news/hong-kong/politics/article/3014695/sea-black-hong-kong-will-march-against-suspended (visited on 04/29/2022).

Rosenblatt, Frank (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386.

Ruder, Sebastian (2016). "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747*.

Saab, Rim, Tausch, Nicole, Spears, Russell, and Cheung, Wing-Yee (Sept. 2015). "Acting in solidarity: Testing an extended dual pathway model of collective action by bystander group members". In: *British Journal of Social Psychology* 54.3, pp. 539–560. ISSN: 01446665. DOI: 10.1111/bjso.12095. URL: https://onlinelibrary.wiley.com/doi/10.1111/bjso.12095 (visited on 04/18/2022).

Sanfey, Alan G., Rilling, James K., Aronson, Jessica A., Nystrom, Leigh E., and Cohen, Jonathan D. (June 13, 2003). "The neural basis of economic decision-making in the Ultimatum Game". In: *Science (New York, N.Y.)* 300.5626, pp. 1755–1758. ISSN: 1095-9203. DOI: 10.1126/science.1082976.

Sanh, Victor, Debut, Lysandre, Chaumond, Julien, and Wolf, Thomas (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *arXiv preprint arXiv:1910.01108*.

Saravia, Elvis, Liu, Hsien-Chi Toby, Huang, Yen-Hao, Wu, Junlin, and Chen, Yi-Shin (2018). "Carer: Contextualized affect representations for emotion recognition". In: *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 3687–3697.

Sass, Louis A. and Parnas, Josef (2003). "Schizophrenia, consciousness, and the self". In: *Schizophrenia Bulletin* 29.3, pp. 427–444. ISSN: 0586-7614. DOI: 10.1093/oxfordjournals.schbul.a007017.

Sattelberg, William (2021). *The Demographics of Reddit: Who Uses the Site?* Alphr. URL: https://www.alphr.com/demographics-reddit/ (visited on 11/25/2021).

Saurkar, Anand V., Pathare, Kedar G., and Gode, Shweta A. (Apr. 30, 2018). "An Overview On Web Scraping Techniques And Tools". In: *International Journal on Future Revolution in Computer Science & Communication Engineering* 4.4, pp. 363–367. ISSN: 2454-4248. URL: http://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1529 (visited on 11/26/2021).

Sen, Anirban, Rudra, Koustav, and Ghosh, Saptarshi (Jan. 2015). "Extracting situational awareness from microblogs during disaster events". In: *2015 7th International Conference on Communication Systems and Networks (COMSNETS)*. 2015 7th International Conference on Communication Systems and Networks (COMSNETS). ISSN: 2155-2509, pp. 1–6. DOI: 10.1109/COMSNETS.2015.7098720.

Serde (2021). *Overview · Serde*. URL: https://serde.rs/ (visited on 11/26/2021).

Shannon, Claude Elwood (1948). "A mathematical theory of communication". In: *The Bell system technical journal* 27.3, pp. 379–423.

Shepherd, Lee, Fasoli, Fabio, Pereira, Andrea, and Branscombe, Nyla R (2018). "The role of threat, emotions, and prejudice in promoting collective action against immigrant groups". In: *European Journal of Social Psychology* 48.4, pp. 447–459.

Shin, Minchul, Seo, Ju-Hwan, and Kwon, Dong-Soo (2017). "Face image-based age and gender estimation with consideration of ethnic difference". In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 567–572.

Shorten, Connor and Khoshgoftaar, Taghi M (2019). "A survey on image data augmentation for deep learning". In: *Journal of big data* 6.1, pp. 1–48.

Shorten, Connor, Khoshgoftaar, Taghi M, and Furht, Borko (2021). "Text data augmentation for deep learning". In: *Journal of big Data* 8.1, pp. 1–34.

Simonsohn, Uri, Simmons, Joseph P, and Nelson, Leif D (2020). "Specification curve analysis". In: *Nature Human Behaviour* 4.11, pp. 1208–1214.

Simpson, Eleanor H. and Balsam, Peter D. (2016). "The Behavioral Neuroscience of Motivation: An Overview of Concepts, Measures, and Translational Applications". In: *Current Topics in Behavioral Neurosciences* 27, pp. 1–12. ISSN: 1866-3370. DOI: 10.1007/7854_2015_402.

Singh, Gurinder, Kumar, Bhawna, Gaur, Loveleen, and Tyagi, Akriti (2019). "Comparison between multinomial and Bernoulli naive Bayes for text classification". In: *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*. IEEE, pp. 593–596.

Siu, Jasmine (Apr. 13, 2019). *Grisly details emerge over student's killing of girlfriend in Taiwan*. South China Morning Post. URL: https://www.scmp.com/news/hong-kong/law-and-crime/article/3005990/body-folded-suitcase-gruesome-details-emerge-hong-kong (visited on 04/27/2022).

Smith, Elaine M, González, Roberto, and Frigolett, Cristián (2021). "Understanding Change in Social-Movement Participation: The Roles of Social Norms and Group Efficacy". In: *Political Psychology* 42.6, pp. 1037–1051.

Smith, Eliot R and Conrey, Frederica R (2007). "Agent-based modeling: A new approach for theory building in social psychology". In: *Personality and social psychology review* 11.1, pp. 87–104.

Smola, Alex J and Schölkopf, Bernhard (2004). "A tutorial on support vector regression". In: *Statistics and computing* 14.3, pp. 199–222.

Spaiser, Viktoria, Chadefaux, Thomas, Donnay, Karsten, Russmann, Fabian, and Helbing, Dirk (Apr. 3, 2017). "Communication power struggles on social media: A case study of the 2011–12 Russian protests". In: *Journal of Information Technology & Politics* 14.2, pp. 132–153. ISSN: 1933-1681. DOI: 10.1080/19331681.2017.1308288. URL: https://doi.org/10.1080/19331681.2017.1308288 (visited on 04/18/2022).

Spaiser, Viktoria, Nisbett, Nicole, and Stefan, Cristina G (2022). ""How dare you?"—The normative challenge posed by Fridays for Future". In: *PLOS Climate* 1.10, e0000053.

Standard, The (July 22, 2019). *Junius Ho accused of supporting Yuen Long mob*. The Standard. URL: https://www.thestandard.com.hk/breaking-news/section/3/131702/Junius-Ho-accused-of-supporting-Yuen-Long-mob (visited on 05/01/2022).

Stott, C. and Reicher, S. (May 1998). "How Conflict Escalates: The Inter-Group Dynamics of Collective Football Crowd 'Violence'". In: *Sociology* 32.2, pp. 353–377. ISSN: 0038-0385, 1469-8684. DOI: 10.1177/0038038598032002007. URL: http://journals.sagepub.com/doi/10.1177/0038038598032002007 (visited on 04/18/2022).

Stott, Clifford and Drury, John (1999). "The Inter-Group Dynamics of Empowerment: A Social Identity Model". In: *Transforming Politics: Power and Resistance*. Ed. by Paul Bagguley and Jeff Hearn. London: Palgrave Macmillan UK, pp. 32–45. ISBN: 9781349274291. DOI: 10.1007/978-1-349-27429-1_3. URL: https://doi.org/10.1007/978-1-349-27429-1_3 (visited on 04/18/2022).

Swann, William B. and Bosson, Jennifer K. (June 30, 2010). "Self and Identity". In: *Handbook of Social Psychology*. Ed. by Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey. Hoboken, NJ, USA: John Wiley & Sons, Inc., socpsy001016. ISBN: 9780470561119. DOI: 10.1002/9780470561119.socpsy001016. URL: https://onlinelibrary.wiley.com/doi/10.1002/9780470561119.socpsy001016 (visited on 04/18/2022).

Tajfel, Henri (Apr. 1974). "Social identity and intergroup behaviour". In: *Social Science Information* 13.2, pp. 65–93. ISSN: 0539-0184, 1461-7412. DOI: 10.1177/053901847401300204.

URL: http://journals.sagepub.com/doi/10.1177/053901847401300204 (visited on 04/18/2022).

– ed. (1978). *Differentiation between social groups: studies in the social psychology of intergroup relations*. European monographs in social psychology 14. London ; New York: Published in cooperation with European Association of Experimental Social Psychology by Academic Press. 474 pp. ISBN: 9780126825503.

Tajfel, Henri, Turner, John C, Austin, William G, and Worchel, Stephen (1979). "An integrative theory of intergroup conflict". In: *Organizational identity: A reader* 56.65, pp. 9780203505984–16.

Tajfel, Henri and Turner, John C. (2004). "The Social Identity Theory of Intergroup Behavior". In: *Political Psychology*. Psychology Press. ISBN: 9780203505984.

Tarrow, Sidney (1998). *Power in Movement: Social Movements and Contentious Politics*. 2nd ed. Cambridge Studies in Comparative Politics. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511813245. URL: https://www.cambridge.org/core/books/power-in-movement/E9FC85E59075F0705549710D6A8BD858 (visited on 04/18/2022).

Tausch, Nicole and Becker, Julia C (2013). "Emotional reactions to success and failure of collective action as predictors of future action intentions: A longitudinal investigation in the context of student protests in Germany". In: *British Journal of Social Psychology* 52.3, pp. 525–542.

Tausch, Nicole, Becker, Julia C, Spears, Russell, Christ, Oliver, Saab, Rim, Singh, Purnima, and Siddiqui, Roomana N (2011a). "Explaining radical group behavior: Developing emotion and efficacy routes to normative and nonnormative collective action." In: *Journal of personality and social psychology* 101.1, p. 129.

– (July 2011b). "Explaining radical group behavior: Developing emotion and efficacy routes to normative and nonnormative collective action". In: *Journal of Personality and Social Psychology* 101.1, pp. 129–148. ISSN: 1939-1315. DOI: 10.1037/a0022728.

Thelwall, Mike, Buckley, Kevan, Paltoglou, Georgios, Cai, Di, and Kappas, Arvid (2010). "Sentiment strength detection in short informal text". In: *Journal of the American society for information science and technology* 61.12, pp. 2544–2558.

Theocharis, Yannis and Jungherr, Andreas (Mar. 15, 2021). "Computational Social Science and the Study of Political Communication". In: *Political Communication* 38.1, pp. 1–22. ISSN: 1058-4609. DOI: 10.1080/10584609.2020.1833121. URL: https://doi.org/10.1080/10584609.2020.1833121 (visited on 11/24/2021).

Tilly, Charles (Jan. 31, 2004). *Social Movements, 1768–2004*. New York: Routledge. 204 pp. ISBN: 9781315632063. DOI: 10.4324/9781315632063.

Toda, Hiro Y and Phillips, Peter CB (1994). "Vector autoregression and causality: a theoretical overview and simulation study". In: *Econometric reviews* 13.2, pp. 259–285.

Toda, Hiroyuki (1991). *Vector autoregression and causality*. Yale University.

Tong, Simon and Koller, Daphne (Mar. 1, 2002). "Support vector machine active learning with applications to text classification". In: *The Journal of Machine Learning Research* 2, pp. 45–66. ISSN: 1532-4435. DOI: 10.1162/153244302760185243. URL: https://doi.org/10.1162/153244302760185243 (visited on 12/05/2021).

Tripathi, Anjaneya (2022). *Emotion Classification NLP*. URL: https://kaggle.com/anjaneyatripathi/emotion-classification-nlp (visited on 03/12/2022).

Tropp, Linda R and Brown, Amy C (2004). "What benefits the group can also benefit the individual: Group-enhancing and individual-enhancing motives for collective action". In: *Group processes & intergroup relations* 7.3, pp. 267–282.

Twitter (2021a). *Consuming streaming data*. URL: https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data (visited on 11/26/2021).

– (2021b). *Tweet object | Docs | Twitter Developer Platform*. URL: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet (visited on 11/25/2021).

– (2021c). *Twitter API for Academic Research | Products*. URL: https://developer.twitter.com/en/products/twitter-api/academic-research (visited on 11/25/2021).

Van Laer, Jeroen and Van Aelst, Peter (Dec. 1, 2010). "Internet and Social Movement Action Repertoires". In: *Information, Communication & Society* 13.8, pp. 1146–1171. ISSN: 1369-

118X. DOI: 10.1080/13691181003628307. URL: https://doi.org/10.1080/13691181003628307 (visited on 04/18/2022).

Van Stekelenburg, Jacquelien and Klandermans, Bert (2017). "Individuals in Movements: A Social Psychology of Contention". In: *Handbook of Social Movements Across Disciplines*. Ed. by Conny Roggeband and Bert Klandermans. Cham: Springer International Publishing, pp. 103–139. ISBN: 9783319576480. DOI: 10.1007/978-3-319-57648-0_5. URL: https://doi.org/10.1007/978-3-319-57648-0_5 (visited on 04/18/2022).

Van Stekelenburg, Jacquelien, Klandermans, Bert, and Van Dijk, Wilco W. (Jan. 1, 2011). "Combining motivations and emotion: The motivational dynamics of protest participation". In: *International Journal of Social Psychology* 26.1, pp. 91–104. ISSN: 0213-4748. DOI: 10.1174/021347411794078426. URL: https://doi.org/10.1174/021347411794078426 (visited on 04/18/2022).

Van Zomeren, Martijn, Leach, Colin Wayne, and Spears, Russell (May 2012). "Protesters as "Passionate Economists": A Dynamic Dual Pathway Model of Approach Coping With Collective Disadvantage". In: *Personality and Social Psychology Review* 16.2, pp. 180–199. ISSN: 1088-8683, 1532-7957. DOI: 10.1177/1088868311430835. URL: http://journals.sagepub.com/doi/10.1177/1088868311430835 (visited on 04/18/2022).

Van Zomeren, Martijn, Postmes, Tom, and Spears, Russell (July 2008). "Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives". In: *Psychological Bulletin* 134.4, pp. 504–535. ISSN: 0033-2909. DOI: 10.1037/0033-2909.134.4.504.

Van Zomeren, Martijn, Spears, Russell, Fischer, Agneta H., and Leach, Colin Wayne (Nov. 2004). "Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy". In: *Journal of Personality and Social Psychology* 87.5, pp. 649–664. ISSN: 0022-3514. DOI: 10.1037/0022-3514.87.5.649.

Van Zomeren, Martijn, Spears, Russell, and Leach, Colin Wayne (Dec. 1, 2010). "Experimental evidence for a dual pathway model analysis of coping with the climate crisis". In: *Journal of Environmental Psychology* 30.4, pp. 339–346. ISSN: 0272-4944. DOI: 10.1016/j.

jenvp.2010.02.006. URL: https://www.sciencedirect.com/science/article/pii/S0272494410000265 (visited on 04/18/2022).

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, and Polosukhin, Illia (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc. URL: https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (visited on 12/07/2021).

Verma, Sudha, Vieweg, Sarah, Corvey, William, Palen, Leysia, Martin, James, Palmer, Martha, Schram, Aaron, and Anderson, Kenneth (2011). "Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency". In: *Proceedings of the International AAAI Conference on Web and Social Media* 5.1, pp. 385–392. ISSN: 2334-0770. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14119 (visited on 12/02/2021).

Verweij, Marco and Damasio, Antonio (May 23, 2019). *The Somatic Marker Hypothesis and Political Life*. Oxford Research Encyclopedia of Politics. DOI: 10.1093/acrefore/9780190228637.013.928. URL: https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-928 (visited on 04/18/2022).

Vrieze, Scott I (2012). "Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)." In: *Psychological methods* 17.2, p. 228.

Wali, Esma, Chen, Yan, Mahoney, Christopher, Middleton, Thomas, Babaeianjelodar, Marzieh, Njie, Mariama, and Matthews, Jeanna Neefe (July 11, 2020). "Is Machine Learning Speaking my Language? A Critical Look at the NLP-Pipeline Across 8 Human Languages". In: *arXiv:2007.05872 [cs]*. arXiv: 2007.05872. URL: http://arxiv.org/abs/2007.05872 (visited on 12/02/2021).

Wang, Bo, Tsakalidis, Adam, Liakata, Maria, Zubiaga, Arkaitz, Procter, Rob, and Jensen, Eric (Apr. 2016). "SMILE Twitter Emotion dataset". In: DOI: 10.6084/m9.figshare.3187909.v2. URL: https://figshare.com/articles/dataset/smile_annotations_final_csv/3187909.

Wei, Jason and Zou, Kai (2019). "Eda: Easy data augmentation techniques for boosting performance on text classification tasks". In: *arXiv preprint arXiv:1901.11196*.

Weinreich, Peter (1986). "The operationalisation of identity theory in racial and ethnic relations". In: *Theories of Race and Ethnic Relations*. Ed. by David Mason and John Rex. Comparative Ethnic and Race Relations. Cambridge: Cambridge University Press, pp. 299–320. ISBN: 9780521369398. DOI: 10.1017/CBO9780511557828.016. URL: https://www.cambridge.org/core/books/theories-of-race-and-ethnic-relations/operationalisation-of-identity-theory-in-racial-and-ethnic-relations/EC59B5971595648280FA75720DC25918 (visited on 04/18/2022).

Wright, Stephen C. (Jan. 1, 2003). "Strategic Collective Action: Social Psychology and Social Change". In: *Blackwell Handbook of Social Psychology: Intergroup Processes*. Ed. by Rupert Brown and Samuel L. Gaertner. Oxford, UK: Blackwell Publishers Ltd, pp. 409–430. ISBN: 9780631210627. DOI: 10.1002/9780470693421.ch20. URL: https://onlinelibrary.wiley.com/doi/10.1002/9780470693421.ch20 (visited on 04/18/2022).

Yang, Yiming and Pedersen, Jan O. (July 8, 1997). "A Comparative Study on Feature Selection in Text Categorization". In: *Proceedings of the Fourteenth International Conference on Machine Learning*. ICML '97. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 412–420. ISBN: 9781558604865. (Visited on 12/02/2021).

Zhang, Harry (2004). "The optimality of naive Bayes". In: *Aa* 1.2, p. 3.

Appendix A

Appendix Chapter 4

A.1 Keywords used in Twitter data collection

#ActOnClimate
#AmericanYellowVests
#Bolsonaro
#Brexit
#BrexitBarometer
#CaravanaDeMigrantes
#CaravanaMigrante
#ClimateBreakdown
#ClimateStrike
#ComandoJungla
#EarthStrike
#ExtinctionRebellion
#ExtinctionRebellionIndia
#ExtinctionRebellionUSA
#Fridaysforfuture
#GiletsJaunes
#KlimatStrejk
#MigrantCaravan
#MigrantCaravans
#XRScotland
#XRSnowflakes
#XRlinkup
#YellowVests
#YellowVestsProtests
#YellowVestsUK
#araucania
#bollockstobrexit
#brexit
#camilocatrillanca
#caravanamigrantes
#humansofXR
#mapuche
#stopbrexit
#wallmapu
#weichafe
@EsXrebellion @ExtinctRebelRI
@ExtinctRebelsIE
@ExtinctionI
@ExtinctionMilan
@ExtinctionR
@ExtinctionRDK
@ExtinctionR_DE
@ExtinctionRebe2
@ExtinctionRebe6
@ExtinctionRebel
@ExtinctionT
@ExtinctionX
@Extinctionrmanc
@GiletsJaunesFr
@GiletsJaunesFr_
@GiletsJaunesGo
@JaunesLes
@NLRebellion
@RebellionTwin
@RisingUpUK
@ScotlandXr
@StrikeClimate
@XRBerlin
@XRBrighton

@XRBristol	@XRboston	Bolsonaro	#YouthStrike4Climate
@XRExeter	@XRebellionAus	Novembre2018	#Youth4Climate
@XRLancs	@XRebellionUS	#LeaveMeansLeave	#FridaysForFuture
@XRSeattle	@XtinctionRebel	#LeaveEU	#SchoolStrike4Climate
@XR_Southampton	@_Gilets_Jaunes_	#VoteLeave	
@XR_Sweden	@cambridge_xr	#Remain	
@XR_York	@jairbolsonaro	#BrexitShambles	

A.2 Code Repositories

Data collection code repository: https://github.com/dhvalden/phd

Fully working Toy Multilayer Perceptron repository: https://github.com/dhvalden/toynn

A.3 Example Twitter JSON Object


```
{
  "text": "RT @PostGradProblem: In preparation for the NFL lockout...",
  "truncated": true,
  "in_reply_to_user_id": null,
  "in_reply_to_status_id": null,
  "favorited": false,
  "source": "<a href=\"http://twitter.com/\" rel=\"nofollow\">Twitter for iPhone</a>",
  "in_reply_to_screen_name": null,
  "in_reply_to_status_id_str": null,
  "id_str": "54691802283900928",
  "entities": {
        "user_mentions": [
              {
                    "indices": [
                          3,
                          19
                    ],
                    "screen_name": "PostGradProblem",
                    "id_str": "271572434",
                    "name": "PostGradProblems",
                    "id": 271572434
              }
        ],
        "urls": [ ],
        "hashtags": [ ]
  },
  "contributors": null,
  "retweeted": false,
  "in_reply_to_user_id_str": null,
  "place": null,
  "retweet_count": 4,
  "created_at": "Sun Apr 03 23:48:36 +0000 2011",
  "retweeted_status": {
        "text": "In preparation for the NFL lockout...,
        "truncated": false,
        "in_reply_to_user_id": null,
        "in_reply_to_status_id": null,
```

```
"favorited": false,
"source": "<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">HootSuite</a>",
"in_reply_to_screen_name": null,
"in_reply_to_status_id_str": null,
"id_str": "54640519019642881",
"entities": {
      "user_mentions": [ ],
      "urls": [ ],
      "hashtags": [
            {
                  "text": "PGP",
                  "indices": [
                        130,
                        134
                  ]
            }
      ]
},
"contributors": null,
"retweeted": false,
"in_reply_to_user_id_str": null,
"place": null,
"retweet_count": 4,
"created_at": "Sun Apr 03 20:24:49 +0000 2011",
"user": {
      "notifications": null,
      "profile_use_background_image": true,
      "statuses_count": 31,
      "profile_background_color": "C0DEED",
      "followers_count": 3066,
      "profile_image_url": "http://a2.twimg.com/profile_images/170264/PGP_normal.jpg",
      "listed_count": 6,
      "profile_background_image_url": "http://a3.twimg.com/a/1306/images/themes/theme1/bg.png",
      "description": "",
      "screen_name": "PostGradProblem",
      "default_profile": true,
      "verified": false,
      "time_zone": null,
      "profile_text_color": "333333",
      "is_translator": false,
      "profile_sidebar_fill_color": "DDEEF6",
      "location": "",
      "id_str": "271572434",
      "default_profile_image": false,
      "profile_background_tile": false,
```

```
            "lang": "en",
            "friends_count": 21,
            "protected": false,
            "favourites_count": 0,
            "created_at": "Thu Mar 24 19:45:44 +0000 2011",
            "profile_link_color": "0084B4",
            "name": "PostGradProblems",
            "show_all_inline_media": false,
            "follow_request_sent": null,
            "geo_enabled": false,
            "profile_sidebar_border_color": "C0DEED",
            "url": null,
            "id": 271572434,
            "contributors_enabled": false,
            "following": null,
            "utc_offset": null
      },
      "id": 54640519019642880,
      "coordinates": null,
      "geo": null
},
"user": {
      "notifications": null,
      "profile_use_background_image": true,
      "statuses_count": 351,
      "profile_background_color": "C0DEED",
      "followers_count": 48,
      "profile_image_url": "http://a1.twimg.com/profile_images/455128973/o1_500_normal.jpg",
      "listed_count": 0,
      "profile_background_image_url": "http://a3.twimg.com/a/1300479984/images/themes/theme1/bg.png",
      "description": "watcha doin in my waters?",
      "screen_name": "OldGREG85",
      "default_profile": true,
      "verified": false,
      "time_zone": "Hawaii",
      "profile_text_color": "333333",
      "is_translator": false,
      "profile_sidebar_fill_color": "DDEEF6",
      "location": "Texas",
      "id_str": "80177619",
      "default_profile_image": false,
      "profile_background_tile": false,
      "lang": "en",
      "friends_count": 81,
      "protected": false,
```

```
            "favourites_count": 0,
            "created_at": "Tue Oct 06 01:13:17 +0000 2009",
            "profile_link_color": "0084B4",
            "name": "GG",
            "show_all_inline_media": false,
            "follow_request_sent": null,
            "geo_enabled": false,
            "profile_sidebar_border_color": "C0DEED",
            "url": null,
            "id": 80177619,
            "contributors_enabled": false,
            "following": null,
            "utc_offset": -36000
      },
      "id": 54691802283900930,
      "coordinates": null,
      "geo": null
}
```

Appendix B

Appendix Chapter 5

B.1 Appendix Section 5.1

Note on Transformers training process

As mentioned in section 4.2.4, transformers-based classifiers underwent a slightly different process of training than the other, simpler, classifiers. Transformers, as other deep-learning models, can be retrained with the same data multiple times, with decreased risk of over-fitting. This is known as training *epochs*. In order to evaluate the amount of over-fitting accumulated, after each epoch the training and testing loss -an indicator of inaccuracy of the classifier normalised between 0 and 1- is recorded. Figure B.1 shows the evolution per epoch of the test and training loss for the three transformers-based models used.

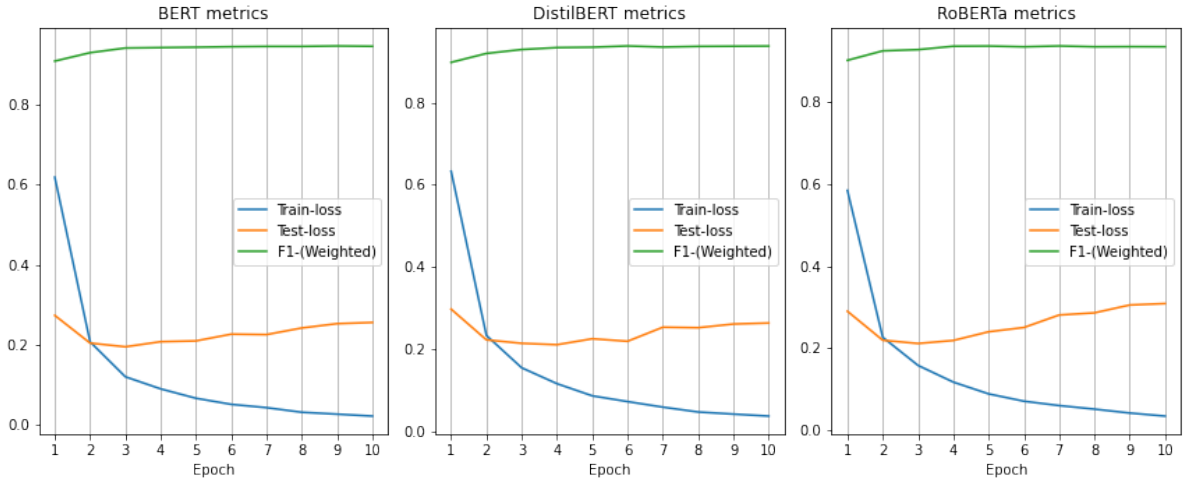

Figure B.1: Evolution of English training and testing loss by epoch of BERT (top), DistilBERT (middle), and RoBERTa (bottom).

As can be seen in figure B.1, training loss shows a monotonic decrease, describing a negative exponential behaviour. Testing loss, on the other hand, has a less smooth behaviour, showing decreasing up until 3 epochs (BERT, RoBERTa) and 4 epochs (DistilBERT). After that, the testing loss value increases. This is a common symptom of over-fitting, which in turn affects

the generalisation capacity of the models. Therefore the checkpoints selected for testing BERT and RoBERTa models where the ones with 3 epochs of training, whereas for DistilBERT with 4 epochs of training.

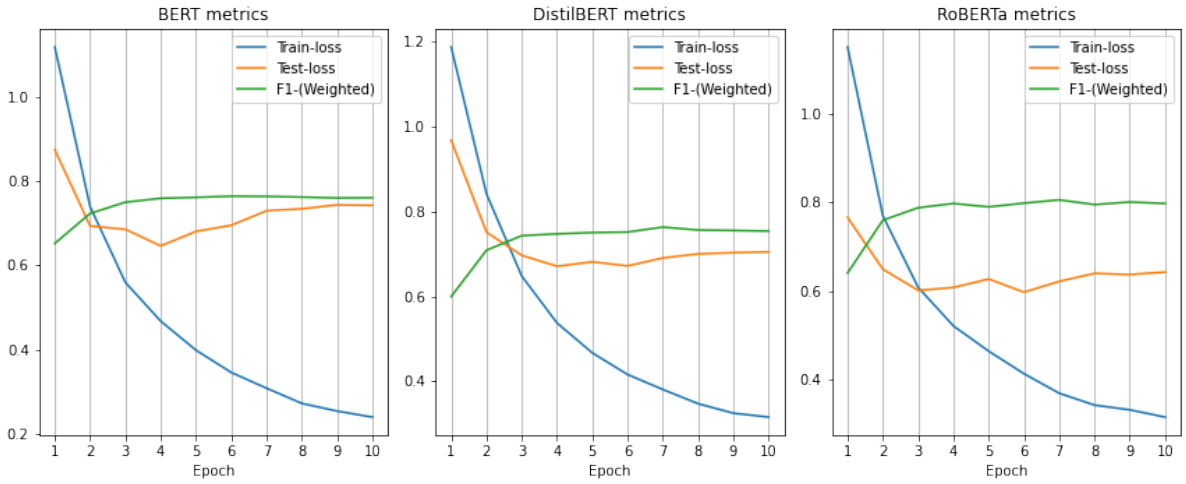

Figure B.2: Evolution of Spanish training and testing loss by epoch of BERT (top), DistilBERT (middle), and RoBERTa (bottom).

Figure B.2 shows the evolution of transformers-based models with the Spanish dataset. Here it can be seen that the behaviours of the training loss curve is again monotonic negative exponential, but the testing loss is decisively less stable. For the evaluation of the models I chose the 4th checkpoint in the case of BERT and DistilBERT, and the 6th checkpoint in the case of RoBERTa.

B.2 Appendix Section 5.2

Average Levenshtein edit-distance Python implementation

```
def av_leve_dist(original: list, new: list):
    average_leve = []
    for i in range(len(original)):
        max_length = max(len(str(original[i])), len(str(new[i])))
        average = edit_distance(str(new[i]), str(original[i])) / max_length
        average_leve.append(average)
    return average_leve
```