Social media forensics applied to assessment of post–critical incident social reaction: The case of the 2017 Manchester Arena terrorist attack

Maxime Bérubé[a, *], Thuc-Uyên Tang[b], Francis Fortin[b], Sefa Ozalp[c], Matthew L. Williams[c], and Pete Burnap[c]

[a] *Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, CAN.*

[b] *School of Criminology, University of Montreal, Montreal, CAN*

[c] *HateLab and Social Data Science Lab, Cardiff University, Cardiff, UK*

* Corresponding author.
E-mail address: maxime.berube2@uqtr.ca (M. Bérubé).
Department of Chemistry, Biochemistry and Physics
3351, boulevard des Forges, CIPP
Trois-Rivières, QC, G9A 5H7
+1819-376-5011

Abstract

The branch of forensic science known as digital forensics is constantly evolving and transforming, reflecting the numerous technological innovations of recent decades. There are, however, continuing issues with the use of digital data, such as the difficulty of handling large-scale collections of text data. As one way of dealing with this problem, we used machine-learning techniques, particularly natural language processing and Latent Dirichlet Allocation (LDA) topic modeling, to create an unsupervised text reduction method that was then used to study social reactions in the aftermath of the 2017 Manchester Arena bombing. Our database was a set of millions of messages posted on Twitter in the first 24 hours after the attack. The findings show that our method improves on the tools presently used by law enforcement and other agencies to monitor social media, particularly following an event that is likely to create widespread social reaction. For example, it makes it possible to track different types of social reactions over time and to identify subevents that have a significant impact on public perceptions.

Keywords :

Digital investigation, Natural language processing, Topic modeling, Terrorism, Intelligence.

Social media forensics applied to assessment of post–critical incident social reaction: The case of the 2017 Manchester Arena terrorist attack

Introduction

Forensic science is constantly evolving and transforming in response to the numerous technological innovations in recent decades. Social media are now frequently used to monitor public behaviours and attitudes, especially following critical public incidents (Williams et al., 2017a, Procter et al., 2013, Innes et al., 2018). Previous studies focused on the use of social media in intelligence and criminal investigations have uncovered misinformation schemes (Burnap et al., 2015), illicit social networks (Yang et al., 2012), or hate speech dissemination processes (Williams and Burnap, 2016, Watanabe et al., 2018), but few studies have focused on the use of digital traces in the context of intelligence-led policing and crisis management.

People use social media to disseminate information or express opinions and increased use, as well as the proliferation of electronically connected devices, means that they are leaving more and more traces of their digital activities. These traces are often referred to as big data and can be used by both social and forensic scientists. Traces can be used to identify criminals

and to gather evidence or can be exploited for other policing tasks, such as investigation, intelligence, and risk assessment.

Big data, despite its richness, poses a number of challenges for researchers and analysts. According to Williams and his colleagues (2017a), big "social" data present six major methodological and technical challenges: volume, variety, velocity, veracity, virtue, and value. This study seeks to address volume challenges by using natural language processing and machine learning techniques, in this case, LDA topic modeling, to deal with text mining and topic identification in analysing the massive number of tweets posted in the aftermath of the 2017 Manchester Arena bombing. Our main objective is to look at how such a method can facilitate the processing of traces from social media and provide insightful information for intelligence and crime prevention purposes.

The United Kingdom has been the target of several terrorist attacks in recent years. Among these is the Manchester Arena bombing, which occurred around 10:30 PM on May 22, 2017 when Salman Abedi, a Briton of Libyan descent, set off an explosive belt in the hall of the arena. The attack left 22 dead and 116 injured, including children. Shortly after the attack, the Islamic State claimed responsibility on social media. The nature of the attack and the large number of injuries and damage elicited reactions from many Twitter users.

The paper is organized as follows: the background section provides information about digital traces, text mining, and natural language processing. Related works are then discussed, including a literature review of online social reaction following a terrorist incident that makes it possible to better understand the context in which the current study took place and introduces the section describing the data and method used. The section on results identifies our main findings and discusses them more generally in order to show the benefits of this method of analysis. Finally, we conclude by highlighting the contributions of this work, as well as the limitations that need to be considered in future studies.

Background

The advent of social media has transformed societal practices as well as those of law enforcement. According to Rainie and Wellman (2012), the popularity of social media has created a new social system, changing interactions between individuals, on the Internet, and in the mobility of technologies. This section addresses recent transformations in the digital environment, particularly the recent digital transformation in forensic science and modern text mining and natural language processing techniques. We focus on the practical aspect of these new methodological approaches.

Digital transformations in forensic science

Forensic science, commonly understood as the study of traces as a way to shed light on past events that affect security (McKemmish, 1999, Fortin, 2014, Crispino, 2006), has evolved over the past thirty years. Traditionally, physical, chemical, and biological traces were the focus of interest but new forms of traces are now also being studied. These new traces, described as "digital", can be generated directly in the commission of a crime or caused by the very nature of the multiple technological innovations that have become part of modern societies (Weyermann et al., 2014). Several digital tools provide for the production and storage of data for reference purposes, making it possible to monitor and learn about several human activities and behaviours. Due to their ubiquity, these digital sensors shed light not only on cybercrime cases but also on traditional crimes, which also leave traces in digital space that can be exploited for investigation and intelligence purposes. Thus, beyond the appearance of new traces that can be studied in forensic science, the arrival of digital traces constitutes an important paradigm shift in this discipline (Casey et al., 2019).

According to Roux and his colleagues (2012) the forensic science discipline should not be divided into sub-branches, but one can generally distinguish at least five different sources of digital traces (Karabiyik, 2015, Resendez et al., 2010):

(1) Computer,

(2) Mobile device,

(3) Network,

(4) Database,

(5) Internet and social media.

From these sources, many types of traces can be exploited, such as text, symbols and signs, imagery, signals, files and folders, geolocation, or behavioral data. In this study, we focus on social media traces, and more specifically text data. As the popularity of social media increases, it leaves an invaluable amount of digital data that can be analyzed. Traces left in digital space, generated by Internet users through their connected devices, provide a gigantic amount of digital data that we often associate with the term "big data", (Roxin and Bouchereau, 2017). Current technology makes it possible to maximize the computing and algorithmic power necessary for the analysis and comparison of such large-scale amounts of data. In particular, it is possible to use the data to identify trends in the economic, social, technical, or legal spheres (Boyd and Crawford, 2012). The advent of social media has therefore transformed societal practices as well as those of law enforcement. According to Rainie and Wellman (2012), the popularity of social media also creates a certain social system. Indeed, changes have occurred in the form of interactions between individuals, the Internet and the mobility of technologies. It is, therefore, all the more important to take into account interactions in the digital environment and to see how these can contribute to

security efforts. One way to do this is to look at the massive amount of textual data left by social media users in the course of such interactions.

Text mining and natural language processing

Several methods have been used to study traces from social media from a forensic perspective. As the amount of data is often too large for a traditional qualitative analysis, computational methods of network and content analysis have been used, depending on the research objectives (Gitari et al., 2015, Watanabe et al., 2018). In some cases, content analysis can also be used for data reduction to make subsequent qualitative analysis possible (Grimmer and Stewart, 2013). To do this, various automated content analysis methods are used to classify textual data. There are two major families for this type of analysis, depending on whether particular categories are being investigated or are yet to be determined.

In either case, analyzing large-scale text collections usually involves the prior application of natural language processing, which combines linguistics, artificial intelligence, and computer science to ensure that the text is comprehensible for computers (Nadkarni et al., 2011). Because human language is very complex, it is necessary to standardize the text before analyzing it, even if such standardization risks introducing interpretation bias. In addition, particularly in the case of social media, the text can contain different characters, writing formats, misspellings, emoticons, and so on, which must be corrected in order to optimize results. There are several pre-

processing techniques for this (Manning and Schütze, 1999). To make text collections more understandable by computers, they should have as few abbreviations and misspellings as possible. It is preferable to convert abbreviations and numbers into words, and to correct extended words (i.e., "yesssssss!!"). In addition, in order to increase algorithm accuracy the text can be cleaned of certain special symbols, punctuation, emojis, carriage returns, tabs, and stop words. Stop words are words that one does not wish to take into account in the analysis, such as determinants (i.e., the, a, an, of, in), specific structural terms (i.e., RT, retweet), and other words that introduce excessive frequency and classification biases. After these corrections are made, tokenization can be undertaken to split the content of the collections into elementary lexical units. This step separates sentences into sequences of words. Then, if the classification categories are known, dictionary methods or supervised methods can be used (Grimmer and Stewart, 2013). Dictionary methods involve comparing the text with words in different dictionaries to find those associated with the desired classification categories. The document is then classified according to the frequency of occurrence of words corresponding to those in the dictionaries. Supervised methods usually "begin with human hand coding of documents into a predetermined set of categories. The human hand coding is then used to train, or supervise statistical models" that are used to classify the remaining documents (Grimmer and Stewart, 2013, p.269).

When the categories are unknown and therefore need to be discovered, the analyst or researcher can rely on artificial intelligence for various clustering techniques (Grimmer and Stewart, 2013). One of the main techniques in this regard is topic modeling, which extracts latent topics from text collections (Blei, 2012). As the topics are not known in advance, an unsupervised machine-learning process needs to look at the statistical distribution of topics among the documents. In order to do so, many topic modeling algorithms exist, such as Latent Semantic Analysis (LSA), Hierarchical Dirichlet Process (HDP), Non-negative Matrix factorization (NMF), and Latent Dirichlet Allocation (LDA). Each of them has limitations, especially with respect to the analysis of short messages and unconventional data sets such as tweets (Tang et al., 2014). However, research have also shown that a large volume of data and the fact that tweets are generally very focussed and hardly ever discuss more than 1 or 2 topics make it easier to process and increase the validity of the results (Hong and Davidson, 2010, Zhao et al., 2011). Because LDA is widely documented in the literature and has shown great results in many research on Twitter data, we opted for the latter (Weng et al., 2010, Hong and Davidson, 2010, Gerber, 2014, Ristea et al., 2018). The LDA algorithm can create different topic models using probabilities to determine the number of topics and the keywords associated with them (Blei et al., 2003). “To do this, LDA assumes that each document in a collection is about some set of topics, but that these topics are

distributed unevenly throughout the documents. The topic structure itself is the hidden variable that needs to be derived based on the observed variables, which are the words in the document" (Squire, 2016, p.204). The optimal number of topics is then evaluated using the coherence score for each of the models (O'Callaghan et al., 2015, Syed and Spruit, 2017, Stevens et al., 2012).

Several studies have demonstrated that that combination of traces from social media and topic modeling can help better understand, predict, and prevent crime. The next section looks at related work in this field.

Related work

Because data from social media are easy to access and are generated in real-time, they are increasingly used in forensic science and criminology (Chan and Bennett Moses, 2016). For example, the predictive use of computational models or algorithms can help guide police strategies and decisions in the justice system (Berk and Bleich, 2013). Gerber (2014) integrated specific linguistic features of GPS-tagged tweets in a major city in the United States with a kernel density estimation (KDE) of crime distribution, which allowed him to improve the crime hotspot prediction model for 19 of the 25 types of crime. Similarly, Ristea and her colleagues (2018) use LDA and geolocation to estimate the spatial distribution of crimes based of tweets that were sent from a specific area. Burnap and Williams (2016) have created different models that can help police and political decisionmakers evaluate social media data, particularly in classifying

messages on Twitter expressing cyber hate based on ethnic, sexual preference, and disability characteristics.

However, very few studies have attempted to determine what we can learn from social media interactions in terms of critical post-incident social response for intelligence and crime prevention purposes. One exception is work by Williams et al. (2017a), who define each Twitter user as a sensor of offline phenomenon that can send information related to social and physical disorder from one of four perspectives: victim, witness, observer, and perpetrator (Williams et al., 2017a). Various stakeholders can, among other things, disclose temporal and spatial information about the evolution of a particular situation. Similarly, Awan and Zempi (2017) suggest that online social reaction has a direct relationship with what can be observed offline. Twitter allows analysis of social mood and emotions in near real-time. Given this, forensics of social media data is crucial for investigation, intelligence, and risk assessment, as well as for monitoring public reaction to critical situations, such as a terrorist attack.

Online social reaction following a terrorist incident

The Twitter platform is particularly useful for observing the reactions of populations experiencing a crisis (Cheong and Cheong, 2011, Simon et al., 2015). People are increasingly likely to turn to social media to express themselves and to attempt to better understand the situations in which they find themselves (Ross et al., 2018). The online social reaction following a

terrorist attack is, however, very diverse, ranging from collective solidarity and the offer of support for the families and relatives of the victims (Innes et al., 2018, Magdy et al., 2015) to the spread of hate speech towards the perpetrators or their related community (Awan and Zempi, 2016, Kaakinen et al., 2018, Magdy et al., 2015, Williams and Burnap, 2016). Williams and Burnap (2016) studied tweets that were sent following the Woolwich terrorist attack in 2013. Using Cohen's (1972) typology, they identified three phases of social reaction. First is the *impact phase*, which corresponds to the moment when the triggering event – a terrorist attack in this case – occurs, generating immediate dissemination of information about the deaths, injuries, and destruction that result. During this phase, which generally takes shape in the first hours after the attack, "the social media reaction may affect the nature, extent, and development of deviant activity from spectators" (Williams and Burnap, 2016, p.216). The number of social media communications are likely to skyrocket and rumours spread. Social media also opens the door for the spread of hate and collective mobilization against groups that share the same characteristics as the attackers, in this case, Muslim communities. However, they found such hate speech in only 1% of tweets in their data base. As well, tweets containing hate speech were 45% less likely to be retweeted compared to other types of tweets (Williams and Burnap, 2016). Second is the *inventory phase*, "during which those exposed to the disaster begin to form a preliminary picture of what has happened and of

their own condition" (Cohen, 1972, p.12). During the first two stages, Twitter users try to understand what is going on, often making comparisons to similar events and assumptions about the identity and motives of the perpetrators (Williams and Burnap, 2016). Finally, they identified a *reaction phase*, in which social media communications focused more on the issues raised by the event, such as national security and Islamophobia, rather than on the attack as such.

Other studies have also shown that the Woolwich attack sparked a range of online reactions. Innes et al. (2018) qualitatively analyzed tweets and blog texts related to this event and identified ten components of the social reaction, which they call the ten "Rs" – reporting, requesting, rumouring, responding, recruiting, risking, retaliating, remembering, reheating, and resiliencing – and which can be well integrated with Williams and Burnap's (2016) phases. They found many cases where witnesses to the event reported what they had just seen or heard. This was followed by requests for information or updates about the attack. Rumouring began to occur within the first hours after the attack. As details emerged about the Woolwich attack, Internet users responded with emotional, cognitive, and behavioural comments. The authors also note that groups used these channels to recruit and that users employed them to amplify or reduce public perception of the level of risk associated with the event. Other messages encouraged and legitimized violent acts in retaliation, while some

publications commemorated the victims in order to build a collective memory by which to remember them. Associations between the attack and other events were also made in order to increase tension or blame a certain group. Finally, they identified prosocial messages showing resiliency, solidarity, and unity (Innes et al., 2018).

As we noted, social media have the potential to play an amplifying role in the social reaction. A high degree of identification with the victims and geographic proximity to an attack can also provoke a greater reaction (Conejero and Etxebarria, 2007, Legewie, 2013). Pro-social reactions can also result from an attack, such as greater social cohesion and increased selflessness (McCormack and McKellar, 2015). However, minorities are frequently stigmatized (Hanes and Machin, 2014, Legewie, 2013), because people tend to favour members of their own social group to the detriment of others (Kaakinen et al., 2018). Events that trigger reactions on Twitter are often associated with an increase in the strength of negative feelings expressed (Thelwall et al., 2011). There are also several pro-social reactions that manifest themselves online, such as calls for calm, collective solidarity, resilience, and the defense of Islam (Innes et al., 2018, Magdy et al., 2015). For example, following the attacks that took place in Paris in November 2015, there were greater numbers of tweets defending Muslims and Islam than attacking them (Magdy et al., 2015). These positive tweets offered support to Parisians and attempted to dissociate the Muslim religion from

terrorism. Online social response to a terrorist attack can be as diverse and complex as that offline.

To our knowledge, ours is the first study to apply LDA topic modeling techniques to analyzing social reactions on Twitter following the 2017 Manchester Arena bombing. The next section describes the data and methodology used in the current study.

Data and methodology

The main objective of this study was to apply natural language processing and machine learning techniques to a massive number of tweets published after the Manchester bombing. These tools were used to determine the evolution of social reaction on Twitter as well as the nature of the published comments. This section presents the methodology used to achieve the research objective.

Data

Data were collected using Twitter API and ESRC COSMOS Open Data Analytics software (http://socialdatalab.net/cosmos). Data collection started at 11:49 PM (GST+1) on May 22, 2017, about 1 hour and 19 minutes after the attack, and ended 24 hours later. Since planning for a security response must be done as quickly as possible, we chose to limit our study to one day after the event in order to see what types of reactions occur relatively soon after the incident. All tweets that use the terms “Manchester” or “Ariana” were

collected during this period and were saved every hour in a separate JSON file. Each file contains an hour of tweets, except for first, which lasts for 10 minutes between 11:49 PM and 11:59 PM. Each tweet file contains the date and time of the creation of the tweet, username, retweets, number of likes, hashtags, hyperlinks, and replies containing the terms "Manchester" and "Ariana" combined with other words. For analysis purposes, special characters, URLs, carriage returns, tabs, special symbols, emojis, user names, retweets, "RT" words, and # were deleted. All words were converted to lowercase letters, and all tweets in a language other than English were deleted. In order to analyze the data more precisely, the period studied extended from 11:49 PM on May 22, 2017, to 10:59 PM on May 23, 2017, meaning that 24 time slots were analyzed, with a total of 4,200,444 tweets.

We chose to use Twitter data for many reasons. First, the majority of research done on online social reaction following a large-scale event uses data from this platform (Cheong and Cheong, 2011, Innes et al., 2018, Williams and Burnap, 2016) as messages posted on Twitter are easily accessible, since the majority of them are public (Burnap et al., 2015). (This is not the case with Facebook, for instance.) In addition, since Twitter is a forum where people share their opinions about daily life, it is an excellent source for information on the perceptions and behaviours of the population (Cheong and Cheong, 2011, Innes et al., 2018, Yates and Paquette, 2011). Indeed,

according to Yates and Paquette (2011), Twitter is also a news service, where information circulates quickly.

It is worth noting that special attention has been paid to the ethical framework surrounding the use of Twitter data. As Williams et al. (2017b) suggest, "where the content to be quoted is not sensitive and the user is not vulnerable, researchers may be satisfied with opt-out consent" (p. 1162). Thus, the information relating to media organizations' accounts was not removed. However, we redacted the usernames of personal accounts for which consent has not been obtained (i.e., account deleted at time of writing).

Method

The data were analyzed using multiple Python 3.6 libraries. To convert the JSON data string into an array and to clean up the tweets, Pandas and Numpy libraries were used. Following this, the tokenization process was carried out using SpaCy, while the NLTK library was used for the natural language processing steps, such as the removal of stop words, to which we added the terms "Manchester" and "Ariana". The NLTK library also served to apply a "bag-of-words" approach to the corpus. This process assigns a numerical identifier to each of the words in the corpus – creating a reference dictionary – and then assigns a frequency to each identifier for each corpus analyzed, in this case each of the 24 time periods (Jindal et al., 2015). Once the data were suitable for the LDA topic modeling process, we used the Gensim library to perform coherence score tests to determine the optimal

number of topics for each hour. We were then able to generate one model per hour and to visualize its content with the pyLDAvis library (see Fig. 1).

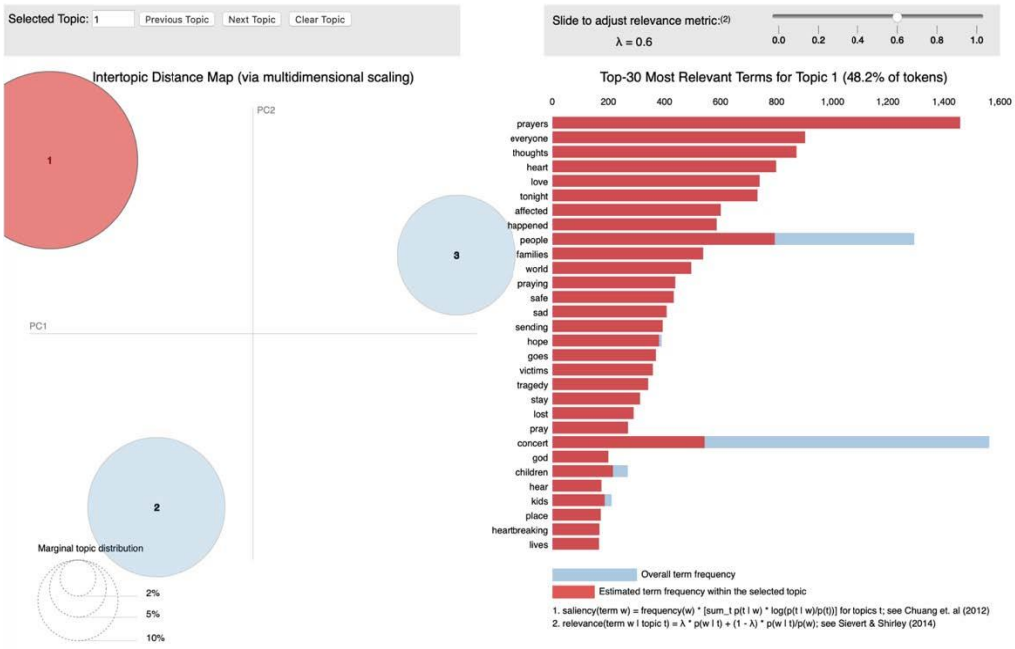


Fig. 1. LDA topic modeling visualization with pyLDAvis

The left side of the interactive visualization makes it possible to locate the number of topics, their similarity, and their prevalence. Topics were created using word frequencies in documents – from the bag-of-words – and a probability calculation was made as to the chances that a word is part of a topic and that this topic is part of a document. The iteration of these random distributions produces topics with words that frequently appear together. Therefore, the same word can be used for several different topics if it is used in different contexts. In this example, there are three separate topics,

represented by the three circles. The size of the circle represents the prevalence of the theme; the more distant the circles, the more distinct the themes. When two circles overlap, it means that the two topics in question are similar and share common keywords. The right side presents a histogram that sorts the main keywords by prevalence among the corpus or within a topic, and also indicates to what extent each keyword is exclusive to a topic. This method allowed us to identify the dominant topics in the tweets in each time slot after the attack and to compare them chronologically. The results section highlights a series of events that caused significant changes in the nature of the tweets and were identified using this methodology.

Limitations

Despite the relevant findings of this study, some limitations must be mentioned. First, it is well known that social media data have biases related to the representativeness and trustworthiness of the studied population (Edwards et al., 2013). It is therefore possible that our data represent the reaction of only specific segments of the population, rather than the population as a whole. Second, the data were arbitrarily split into one-hour periods, which may not correspond to times when topics changed. However, this is a necessary bias and the division still allows for a very fine analysis of the data. A third bias stems from our choice of analysis technique. LDA topic modeling implies a certain degree of subjectivity in identification of topics, since they were not named by the computer (Grimmer and Stewart, 2013). In

order to limit that bias, an interrater reliability process was conducted among the authors to ensure their validity, where authors agreed on topics identified.

Results

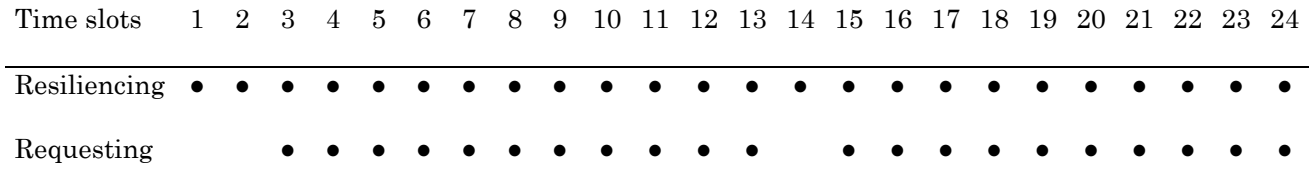
The analysis was performed by creating 24 models, one for each hour. For each model, the number of latent topics per hour according to their best coherence score varies between 2 and 6 (mean=2.79). The topics were generated by the algorithm, and then analyzed qualitatively. We associated each generated topic to six of the categories in Innes et al. (2018) for the entire period studied: categories used were resiliencing, reporting, requesting, responding, remembering, and reheating (see Fig. 2). Figure 2 also shows examples of keywords associated with each concept.

Topic	Description	Example of keywords
Resiliencing	Tweets of sympathy, solidarity, resilience, and tribute to victims	Time slot #4, topic 1: prayers, tonight, love, affected, thoughts, everyone, sending
Requesting	Tweets formulating requests, mainly to find missing people	Time slot #3, topic 1: please, anyone, help, missing, explosion, concert, contact
Reporting	Tweets giving information on the event	Time slot #5, topic 1: explosion, injured, dead,

		moment, video, concert, police, shows, time
Responding	Tweets containing emotional, cognitive and behavioural verbalizations	Time slot #6, topic 3: world, live, make, nothing, let, sad, music, believe, wanted, anything, positivity
Remembering	Tweets on organizing a vigil for the remembrance of the victims	Time slot #20, topic 4: time, silence, kids, vigil, attack, terror, square
Reheating	Tweets of associations between the attack and other events	Time slot #16, topic 2: bombing, isis, explosion, marawi, bomb, bangkok, syria

Fig. 2. Topics, descriptions (Innes et al., 2018), and examples of keywords.

To facilitate understanding the evolution of the topics, we also diagrammed them chronologically (see Fig. 3). It is worth noting that a topic can appear more than once in the same time period but in different forms. For example, two reporting topics relating to different subjects may have been detected in the same time slot.



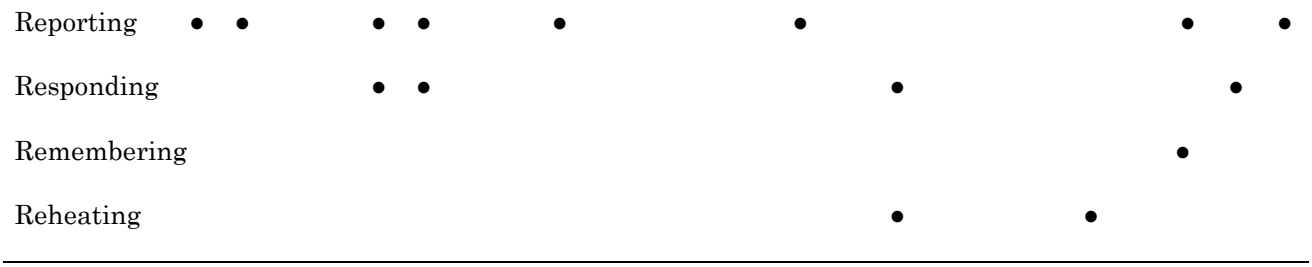


Fig. 3. Topics per time slots.

Our findings show that, for the entire period studied, tweets were largely dominated by resilience messages, requests for help, and reporting of details about the attack. Furthermore, although its format may differ, the topic of resiliencing occurs in each of the per hour models. An example of a different format is when a significant proportion of messages are directed more specifically to the support of American victims and their relatives (mostly between 7:00 PM and 10:59 PM on May 23). Similarly, for the first two time slots, which means before 1:00 AM on May 23, the resiliencing topic includes calls for caution, which then essentially change into messages evoking positive thoughts and prayers for the victims and their relatives:

> "Praying for everyone back in Manchester tonight, stay safe, show no fear." (10:50PM, May 22, 2017).

During these first hours, a significant amount of information on the circumstances of the event is reported. This coincides with a press briefing issued by the Grand Manchester Police (GMP) around midnight. The GMP

confirmed that a major incident had taken place at the Manchester Arena and that there had been numerous deaths and injuries. Shortly after midnight, police said there were 19 confirmed deaths and about 50 injured. This information was quickly and widely shared by Twitter users:

> "RT @CNN: 19 killed, around 50 injured after reports of explosions at Ariana Grande concert in Manchester, British police say…" (00:42AM, May 23, 2017).

Between 1:00 AM and 1:59 AM (time slot #3), about 3 hours after the event, a large number of tweets are requests for help, mainly to find relatives with whom no communication can be established. It is interesting to note that on this topic, the keywords "holiday" and "inn" are salient because a witness to the event, Paula Robinson, was at Victoria train station, located next to the Manchester Arena, when the explosion took place. She saw a group of young people running away from the arena and started running with them. She took them to a nearby hotel and posted a message on Facebook and Twitter to alert their parents that they were at the Holiday Inn:

> "RT @[username redacted]: spread this guys. about 50 children attending the ariana concert in manchester been taken to the holiday inn at vic…" (1:59AM, May 23, 2017).

However, she named the wrong hotel in her tweet and a representative of the Holiday Inn hotel chain messaged that the group of children was not at their hotel. Following this, a significant number of the tweets consist of asking that

users no longer share this tweet as it contains erroneous information, or simply saying that people were still looking for their children and that they were not at the Holiday Inn. Given the large number of tweets shared about this, it is obvious that publication of this information influenced the prevalence of this topic.

After 3:00 AM, a large proportion of resiliencing and reporting tweets were associated with TV host James Corden. He was live on TV when told of the attack and his tweets about the event and honouring the victims were widely shared. The same phenomenon occurs in the morning when the Muslim Council of Britain and Donald Trump condemned the attack. Multiple tweets, for example, echoed comments made by Trump during his press conference:

> "#DonaldTrump - Donald Trump says Manchester Arena attack perpetrators are 'evil losers' https://t.co/rWjiNm2wRD - - - #trump" (8:50AM, May 23, 2017)

Subsequently, several other forms of requesting are seen regarding finding missing persons. They include a significant sharing of requests concerning Olivia Campbell-Hardy, a 15-year-old girl who was at the scene at the time of the attack and was still missing. Tweets at time slot # 13 (11:00 AM to 11:59 AM) are largely composed of messages concerning an evacuation notice for the Manchester Arndale shopping center, launched when police

feared they had discovered a suspicious package. It was later confirmed that the incident was unrelated to the Manchester Arena bombing.

Between 12:00 PM and 12:59 PM, a variety of information about the incident was relayed. Tweets were about the victims, those still missing or found, and the arrest of an individual by the GMP. During this time the police arrested a 23-year-old man possibly linked to the attack, although the identity of the real perpetrator had not yet been revealed.

After 1:00 PM, requests for help resume, especially following the announcement of the death of Saffie Rose Roussos, an 8-year-old girl who had been the subject of previous tweets. A great number of solidarity messages were also sent at this time. Among them were tweets from the Manchester United team, its coach José Mourinho, and the Europa soccer league. The latter also took the opportunity to announce its decision to go ahead with its final match, planned for the next evening in Stockholm. While Ariana Grande's tour seemed to be cancelled and people were worried about refunds, a form of social indignation was expressed in response to this topic:

> “RT @[username redacted]: if ariana cancels the tour i don't want to see anybody bitching... you want your money back?? some parents want their kids” (1:00 PM, May 23, 2017).

From 2:00 PM, shortly after the event was claimed by the Islamic State group, the keyword "marawi" surfaced and led to formation of a topic associated with the concept of reheating. There were confrontations in the

city of Marawi in the Philippines between the Islamic State group fighters and the national army and Filipino internet users mobilized around the Manchester bombing to raise awareness of their situation. The association of the attack with the Islamic State group and with a larger conflict can also serve as a form of risking, as argued by Innes et al. (2018). In this case, it amplified public perception of the Islamist threat to Western populations, as the Manchester Arena bombing was associated with other terrorist incidents:

> "RT @[username redacted]: an explosion in manchester, an isis invasion in marawi, a bomb attack in bangkok, a car bombing in syria all happened in…" (2:00PM, May 23, 2017).

In subsequent hours, the names of the victims and the reactions of the population continue to circulate on Twitter, reflected in the presence of resiliencing and requesting. By the end of the afternoon, sources in the United States circulated information that the person responsible for the explosion was Salman Abedi. This information was later confirmed at a press conference with GMP Chief Ian Hopkins at around 7:00 PM. Hopkins also used the opportunity to ask the people of Manchester to remain united. New details continued to surface about Abedi, in particular that he was born in Manchester and was of Libyan origin as well as that the explosive belt he was wearing contained nuts and bolts and was intended to cause as much injury as possible. In this regard, several tweets report the words of a homeless man who witnessed the event:

> "RT @itvnews: 'We had to pull nails out of children's faces': Steve, a homeless man who was sleeping near #Manchester Arena, rush…" (8:00 PM, May 23, 2017)

This information is then integrated into the topic of reporting. Finally, until the end of the period studied, the topic of resiliencing continues to appear, but we also see tweets that fall under the concept of remembering; at the stroke of 6:00 PM, a large proportion of messages deal with a vigil in memory of the victims held in Albert Square, in front of Manchester Town Hall:

> "RT @DailyMirror: A large crowd is gathered in Albert Square for a vigil for the Manchester bombing victims" (6:01 PM, May 23, 2017)

For the rest of the evening, requests for help also remain, particularly to find a man named Nell Jones, who, like many others, was missing after the attack.

Discussion

Our findings show that LDA unsupervised topic modeling makes it possible to distinguish and monitor various themes over time in large-scale twitter posts. Many concepts suggested by Innes et al. (2018) were present in our corpus and analysis allowed us to identify multiple topics associated with 6 of their 10 "Rs" – resiliencing, reporting, requesting, responding, remembering, and reheating – making it clear that this typology is relevant for study of such data and events. It is important to mention that other

concepts may be present in the corpus but were not of sufficient importance to be defined as a topic by the algorithm. For example, as we saw in the case of attempting to increase the risk perceived by the public by associating the event with Islamist terrorism, some concepts may relate to different interpretations of the same data.

We were also able to find the impact and the inventory phases of social reaction identified by Williams and Burnap (2016). The impact phase began when the Manchester Arena bombing occurred and was followed by immediate reactions from various stakeholders. For instance, the police made announcements in order to keep the population informed about the number of wounded and dead, and the public retweeted these announcements. In the first hours after the attack, various responses from users were found in the generated topics. Among these were various forms of warnings, thoughts, and prayers for the victims and their families, details regarding the attack, and emotional reactions. It is not surprising to see that false information was circulating on Twitter during this phase, such as the case of the group of children who were mistakenly said to have been taken to the Holiday Inn – such rumours usually circulate during the first hours following an attack (Williams and Burnap, 2016, Innes et al., 2018).

The inventory phase started when the population began to get a good idea of what was happening and the damage caused by the incident (Williams and Burnap, 2016). During this period, on the morning of May 23, several

political figures condemned the attack. Relatives were still trying to locate missing persons, as evidenced by the topic of requests for help. Reporting topics were also significant during this period, showing that more and more information was being shared among Twitter users, especially regarding victims and perpetrators. During the inventory phase more precise details generally emerge concerning the victims, those responsible for the attack, and the reasons behind the attack (Williams and Burnap, 2016). It is also during these first phases that people usually formulate hypotheses regarding the identification of those responsible for the attack (Williams and Burnap, 2016). Even before anyone claimed responsibility for the attack, the name of the Islamic State group was circulating in the news and on social media, which confirms that people were trying to understand what was going on.

Finally, the reaction phase began when posts on Twitter focused more on the issues raised by the attack and its root causes, rather than the attack itself (Williams and Burnap, 2016). This includes issues about the effect on national security, Islamophobia, racism, and so forth. This phase was not observed with the topics considered, probably because the period studied was not sufficiently long. In fact, according to Williams and Burnap (2016), the reaction phase in the aftermath of the Woolwich attack started on the fourth day following the incident. In our case, we saw some comments about Islamophobia when President Trump's words were shared, but these were not sufficient to generate a topic.

Conclusion

In today's digital era, forensic science's interest in digital traces has become increasingly important. These traces are invaluable sources of information, although using them effectively poses certain challenges. Acknowledging the undeniable technological changes that have affected the way in which law enforcement and security agencies conduct criminal investigation and intelligence, we hope to show how natural language processing techniques can help in such areas. More specifically, we suggest a method for the identification of themes of interest within thousands of tweets collected over time and in connection with a single event. As part of this study, we focused on the challenge posed by the volume of data accessible on the Internet and social media (Williams et al., 2017a), which is often too large to be processed manually. To overcome this problem, we used machine learning techniques to create an unsupervised text reduction method that significantly reduced the size of our corpus, allowing us to undertake a more qualitative analysis. This method would be very useful for analyzing similar sets of data, greatly facilitating the work of forensic scientists.

Although there are many studies of offline social reactions following a terrorist attack, the range of reactions generated online remains insufficiently explored. Our study attempts to help fill this gap by focusing on the evolution of the social reaction on Twitter following the 2017 Manchester Arena bombing. To understand it, we built a script using multiple Python

libraries and conducted unsupervised LDA topic modeling analysis on 4,200,444 tweets containing the terms “Manchester” and “Ariana” between 11:49 PM on May 22, 2017 and 10:59 PM on May 23, 2017. The results show that it is possible to observe the evolution and diversity of social reactions online.

This study contributes not only to forensic science but to the various fields of study that are intrinsically linked to it. Among other things, and working within a traditional theoretical framework in criminology, it helps understand the different stages of post-critical incident social reaction, especially in an online environment. It also helps understand the social impacts of terrorist events and suggests how different stakeholders should prepare to provide a more effective and efficient response. It provides suggestions for intelligence work by demonstrating that it is possible to identify specific information within a massive amount of text data. The study of online social reactions through tweets can therefore be used to improve the tools used by law enforcement and other agencies to monitor social media, especially in the aftermath of events with the potential to generate social reactions. Indeed, the use of these techniques can allow agencies to get a general idea of the impact of the event very quickly, if not in real time. We have also seen that such a tool is useful to identify information of interest, which can also be very helpful for intelligence analysts and intelligence-led policing (Ratcliffe, 2003) and could provide better information to facilitate

planning prevention strategies and post-incident response, especially through a better understanding of the fears and needs of the concerned citizens.

As mentioned in the method section, our study has its limitations. Among other things, the sample could have been taken over a longer period of time, which would have allowed us to incorporate phases of social reaction that generally occur after the first day following such an event. Also, our method does not allow to fully exploit the thematic results because it ignores the lexicalization of emotions, subjectivity, and polysemy. Despite the qualitative approach that follows the quantitative analysis, which partly overcomes this limit, our tweet extraction method resulted in a loss of the conversational dimension, as well as the link with media elements (images, videos), which certainly could have been relevant elements to consider in the data analysis. Further research should, therefore, focus on gaining a better understanding of the use of these elements of discourse and the linguistic corpora in which these tweets take place. Future studies should also focus on a longer period of time and also try to see how post-critical incident social reaction take shape on other social media platforms. The structural nature of the Twitter environment influences the format and character of posted comments and a more global overview of the situation on various social media is desirable. In the same vein, other natural language processing techniques could be integrated into our models, such as sentiment analysis, which would allow a finer appreciation of the way detected topics are treated.

References

Awan, I. & Zempi, I., 2016. The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. Aggression and Violent Behavior, 27, 1-8.

Awan, I. & Zempi, I., 2017. 'I will blow your face OFF': Virtual and physical world anti-Muslim hate crime. British Journal of Criminology, 57 (2), 362-380.

Berk, R. A. & Bleich, J., 2013. Statistical procedures for forecasting criminal behavior: A comparative assessment. Criminology & Public Policy, 12 (3), 513-544.

Blei, D. M., 2012. Probabilistic topic models. Communications of the ACM, 55 (4), 77-84.

Blei, D. M., Ng, A. Y. & Jordan, M. I., 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

Boyd, D. & Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15 (5), 662-679.

Burnap, P., Gibson, R., Sloan, L., Southern, R. & Williams, M. L., 2015. 140 characters to victory?: Using Twitter to predict the UK 2015 general election. Electoral Studies, 41, 230-233.

Burnap, P. & Williams, M. L., 2016. Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. EPJ Data Science, 5 (11), 1-15.

Casey, E., Ribaux, O. & Roux, C., 2019. The Kodak syndrome: Risks and opportunities created by decentralization of forensic capabilities. Journal of Forensic Sciences, 64 (1), 127-136.

Chan, J. & Bennett Moses, L., 2016. Is big data challenging criminology? Theoretical Criminology, 20 (1), 21-39.

Cheong, F. & Cheong, C. Social media data mining: A social network analysis of tweets during the 2010-2011 Australian floods. PACIS 2011 Proceedings, 2011 Brisbane, Australia. Association for Information Systems.

Cohen, S., 1972. Folk devils and moral panics, London, UK, MacGibbon and Kee Ltd.

Conejero, S. & Etxebarria, I., 2007. The impact of the Madrid bombing on personal emotions, emotional atmosphere and emotional climate. Journal of Social Issues, 63 (2), 273-287.

Crispino, F. 2006. *Le principe de Locard est-il scientifique? Ou analyse de la scientificité des principes fondamentaux de la criminalistique.* PhD Thesis, Université de Lausanne.

Edwards, A., Housley, W., Williams, M. L., Sloan, L. & Williams, M., 2013. Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. International Journal of Social Research Methodology, 16 (3), 245-260.

Fortin, F. 2014. *C'est ma collection mais c'est bien plus que ça: Analyse des processus de collecte et de l'évolution des images dans les collections de pornographie juvénile.* PhD Thesis, Université de Montréal.

Gerber, M. S., 2014. Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, 115-125.

Gitari, N. D., Zuping, Z., Damien, H. & Long, J., 2015. A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering, 10 (4), 215-230.

Grimmer, J. & Stewart, B. M., 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis, 21 (3), 267-297.

Hanes, E. & Machin, S., 2014. Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11. Journal of Contemporary Criminal Justice, 30 (3), 247-267.

Hong, L. & Davidson, B. D. Empirical study of topic modeling in Twitter. Proceedings of the First Workshop on Social Media Analytics, 2010 Washington D.C.: Association for Computing Machinery, 80–88.

Innes, M., Robert, C., Preece, A. & Rogers, D., 2018. Ten "Rs" of social reaction: Using social media to analyse the "post-event" impacts of the murder of Lee Rigby. Terrorism and Political Violence, 30 (3), 454-474.

Jindal, R., Malhotra, R. & Jain, A., 2015. Techniques for text classification: Literature review and current trends. Webology, 12 (2), 1-28.

Kaakinen, M., Oksanen, A. & Rasanen, P., 2018. Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. Computers in Human Behavior, 78, 90-97.

Karabiyik, U. 2015. *Building an intelligent assistant for digital forensics.* PhD Thesis, Florida State University.

Legewie, J., 2013. Terrorist events and attitudes toward immigrants: A natural experiment. American Journal of Sociology, 118 (5), 1199-1245.

Magdy, W., Darwish, K. & Abokhodair, N. Quantifying public response towards Islam on Twitter after Paris attacks. arXiv, 2015.

Manning, C. D. & Schütze, H., 1999. Foundations of statistical natural language processing, Cambridge, MA, The MIT Press.

Mccormack, L. & Mckellar, L., 2015. Adaptative growth following terrorism: Vigilance and anger as facilitators of posttraumatic growth in the aftermath of the Bali bombings. Traumatology, 21 (2), 71-81.

Mckemmish, R., 1999. What is forensic computing? Trends & Issues in Crime and Criminal Justice, 118, 1-6.

Nadkarni, P., Ohno-Machado, L. & Chapman, W. W., 2011. Natural language processing: An introduction. Journal of the American Medical Informatics Association, 18 (5), 544-551.
O'callaghan, D., Green, D., Carthy, J. & Cunningham, P., 2015. An analysis of the coherence of descriptors in topic modeling. Expert Systems with Applications, 42 (13), 5645-5657.
Procter, R., Vis, F. & Voss, A., 2013. Reading the riots on Twitter: Methodological innovation for the analysis of big data. International Journal of Social Research Methodology, 16 (3), 197-214.
Rainie, L. & Wellman, B., 2012. Networked: The new social operating system, Cambridge, MA, The MIT Press.
Ratcliffe, J. H., 2003. Intelligence-led policing, Canberra, AUS, Australian Institute of Criminology.
Resendez, I., Martinez, P. & Abraham, J., 2010. An introduction to digital forensics. ACET Journal of Computer Education and Research, 6 (1).
Ristea, A., Kurland, J., Resch, B., Leitner, M. & Langford, C., 2018. Estimating the spatial distribution of crime events around a football stadium fron georeferences tweets. International Journal of Geo-Information, 7 (2), 1-25.
Ross, B., Potthoff, T., Majchrzak, T. A., Chakraborty, N. R., Ben Lazreg, M. & Stieglitz, S. The diffusion of crisis-related communication on social media: An empirical Analysis of Facebook Reactions. Proceedings of the 51st Hawaii International Conference on System Sciences, 2018 Hawaï, HI. 2525-2534.
Roux, C., Crispino, F. & Ribaux, O., 2012. From forensics to forensic science. Current Issues In Criminal Justice, 24 (1), 7-24.
Roxin, I. & Bouchereau, A., 2017. The ecosystem of the Internet of things. In: Bouhaï, N. & Saleh, I. (eds.) Internet of things: Evolutions and innovations. London, UK: ISTE Ltd, pp. 21-50.
Simon, T., Goldberg, A. & Adini, B., 2015. Socializing in emergencies: A review of the use of social media in emergency situations. International Journal of Information Management, 35 (5), 609-619.
Squire, M., 2016. Mastering data mining with Python: Find patterns hidden in your data, Boston, MA, Safari.
Stevens, K., Kegelmeyer, P., Andrzejewski, D. & Buttler, D. Exploring topic coherence over many models and many topics. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012 Jeju Island, Korea. Association for Computational Linguistics, 952–961.
Syed, S. & Spruit, M. Full-text or abstract? Examining topic coherence scores using Latent Drichlet Allocation. 2017 Conference on data science and advanced analytics, 2017. 165-174.
Tang, J., Meng, Z., Nguyen, X., Mei, Q. & Zhang, M. Understanding the limiting factors of topic modeling via posterior contraction analysis.

Proceedings of the 31st International Conference on Machine Learning, 2014 Beijing, China. Proceedings of Machine Learning Reasearch, 190-198.
Thelwall, M., Buckley, K. & Paltoglou, G., 2011. Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62 (2), 406-418.
Watanabe, H., Bouazizi, M. & Ohtsuki, T., 2018. Hate speech on Twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. IEEE Access, 6, 13825-13835.
Weng, J., Lim, E.-P., Jiang, J. & He, Q. Twitterrank: Finding topic-sensitive influential Twitterers. Proceedings of the Third ACM International Conference on Web Search & Data Mining, 2010 New York, NY. Research Collection School Of Information Systems, 261-270.
Weyermann, C., Jendly, M. & Rossy, Q., 2014. Explorer les intersections entre la science forensique et la criminologie au travers de la temporalité de trois types d'actions de contrôle social. Revue canadienne de criminologie et de police technique, 68 (3), 284-298.
Williams, M. L. & Burnap, P., 2016. Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. British Journal of Criminology, 56 (2), 211-238.
Williams, M. L., Burnap, P. & Sloan, L., 2017a. Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. British Journal of Criminology, 57 (2), 320-340.
Williams, M. L., Burnap, P. & Sloan, L., 2017b. Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. Sociology, 51 (6), 1149–1168.
Yang, C., Harkreader, R., Zhang, J., Shin, S. & Gu, G. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on Twitter. Proceedings of the 21st international conference on World Wide Web, 2012 Lyon, France. Association for Computing Machinery, 71-80.
Yates, D. & Paquette, S., 2011. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. International Journal of Information Management, 31 (1), 6-13.
Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H. & Li, Z., 2011. Comparing Twitter and traditional media using topic models. In: Clough, P., Foley, C., Gurrin, C., Jones, G., Kraaij, W., Lee, H. & Murdock, V. (eds.) Advances in Information Retrieval. Berlin, Germany: Springer, pp. 338-349.