A Conceptual Framework for Social Movements
Analytics for National Security

Pedro Cárdenas\textsuperscript{1}, Georgios Theodoropoulos\textsuperscript{2}, Boguslaw Obara\textsuperscript{1}, Ibad Kureshi\textsuperscript{3}

\textsuperscript{1} Durham University, UK
\textsuperscript{2} Southern University of Science and Technology, CN
\textsuperscript{3} Inlecom Systems, BE
pedro.cardenas-canto@durham.ac.uk, boguslaw.obara@durham.ac.uk
theogeorgios@gmail.com
ibad.kureshi@inlecomsystems.com

Abstract. Social media tools have changed our world due to the way they convey information between individuals; this has led to many social movements either starting on social media or being organised and managed through this medium. At times however, certain human-induced events can trigger Human Security Threats such as Personal Security, Health Security, Economic Security or Political Security. The aim of this paper is to propose a holistic Data Analysis Framework for examining Social Movements and detecting pernicious threats to National Security interests. As a result of this, the proposed framework focuses on three main stages of an event (Detonating Event, Warning Period and Crisis Interpretation) to provide timely additional insights, enabling policy makers, first responders, and authorities to determine the best course of action. The paper also outlines the possible computational techniques utilised to achieve in depth analysis at each stage. The robustness and effectiveness of the framework are demonstrated by dissecting Warning Period scenarios, from real-world events, where the increase of Human Security aspects were key to identifying likely threats to National Security.

Keywords: National Security · Natural Language Processing · Social Movements · Cyberactivism.

1 Introduction

Massive social gatherings and social networks under-pinned by technology are two concepts that walk on the same path, especially when the basic structures or essential norms and values of a social system have been disrupted [14]. As a result of a set of social instability issues, a crisis may be triggered and affect the “homeostasis” or internal balance among those elements that maintain the stability of a state such as the economy, public order, health, environment or even life. Social movements are a clear example of these disruptive events because
people’s behaviour change according to the situation they face and a violent crowd reaction may lead to an instability scenario.

Microblogging websites and services have served as platforms to express ideas as well as to organise and coordinate crowds during a crisis event. Twitter, with over 800 million registered users [24], has seen itself at the centre of several large-scale Social Movements, with individuals conveying their ideas and frustrations within 140 any now 280 characters. Hence, understanding the way Social Movements use microblogs such as Twitter to organise, disseminate ideas, collaborate, coordinate and connect groups or cells of people linked to similar beliefs is, therefore, an essential task to appreciate the evolution of these social events.

There are models that describe how online social movements evolve [16], which parameters describe National Security considerations [21, 25], what computational techniques help to get the private state of individuals, and how to find topics within a data corpus. However, no attention has been paid to create a holistic data analysis framework that links all the above elements and processes it in a timely fashion, to anticipate and detect the core stages of a Social Movement and when the crisis event can affect one or more National Security variables.

The present paper introduces a holistic framework for analysing Social Movements which use Twitter as their primary mean of communication. Our aim is to leverage the capabilities of different computational techniques and aggregating them to understand how these social events evolve and describe a system that can detect whether a social disruption can become a pernicious threat to National Security interests.

The rest of the paper is organised into five sections. Section 2 defines National Security threats, focusing on people as the primary element, and provides a short discussion on the link between social movements and social media. Section 3 offers a high-level description of the proposed framework. Section 4 presents a detailed description of each framework component. Section 5 illustrates the operationalisation of the framework using tweets from the Libyan uprising in 2011, focusing on the earlier stages of the event (referred to as the Warning Period). Section 6 concludes the paper outlining challenges and future work required to realise the proposed framework.

2 National Security in the Social Media Era

Security is a complex concept that has different facets depending on the person or entity in question. At times the different types of security can be at odds with each other. National Security is one of these challenging dimensions. It can be qualified by two main concepts: ensuring the security of the state security; and ensuring the security of its people (Human Security)[21]. [21, 25] make arguments for how Human Security and State Security are mutually supportive.


In the digital era, social media tools have been valuable to spread messages related to those major disasters that have struck a society. Hence, social media platforms can help to identify those human security vulnerabilities that have snowballed into a challenge and required immediate attention.

In the light of the Arab Spring revolutions, the Internet in general and social media networks in particular have gained attention as essential instruments for organising people and communicating ideas and plans. This makes social media the catalyst that enables movements to mobilise hundreds of thousands of individuals in a few hours [22, 16]. Social media facilitates the link between social movements and collective action theory, where individuals share common interests or objectives, and they work as a single unit to accomplish their expectations [13].

One of the ways to analyse the evolution of Social Movements which use virtual platforms is described in a circular flow model proposed by Sandoval et al [16]. Figure 1 demonstrates the links between the four stages of this model, outlined below.

1. **Triggering Event**: This conceives an opportunity in which individuals tend to become active, as a result of a disruptive incident;
2. **Media Response**: This stage considers that the detonating event brings about an instant response supported by a social media platform which allows people and activists to convey ideas, but at the same time works as a natural channel to uncover important events and show them to a domestic or international audience;
3. **Viral Organisation**: Once a detonating event opens a window for individuals to express their political views using a citizen to citizen channel [5], they create online communities where collective ideas of co-production and collaboration are exchanged to reinforce the community engagement;
4. **Physical Response**: The final stage reflects the power of the massive reaction, where protesters tend to organise resistance using different disruptive actions.
3 Conceptual Framework

Figure 2 outlines an iterative cycle that comprises three main stages, forming the core of our proposed model. These steps allow the dissection of the crisis event into core elements that interpret the possible evolution of a National Security instability scenario, and at the end, the results can be used to create a fine-grained strategy (Crisis Scorecard) to deal with the event and determine the best course of action.

This framework takes its root and can be better understood by looking at the medical domain. Within a human health context, diagnosis, detection and interventions are planned using an illness-treatment schema (see Figure 3). In line with this idea, the process begins with the patient assessment, and in a National Security environment, the state plays the patient role. Therefore the illness can be seen as the crisis event that triggered a crowd reaction (Detonating Event). The Diagnosis involves a twofold process; the first step (Warning Period) detects the “symptoms” such as changes in sentiments or opinions over time. When applying this model to Social media, these symptoms activate a computational analysis to identify which National Security variables were affected (Economic Security, Food Security, Health Security, Environmental Security, Personal Security, Communal Security and Political Security).

Once the former analysis reaches a threshold based on domestic National Security Policies, it starts a second step (Crisis Interpretation) which is focused on recognising and analysing other societal characteristics such as violence; coordination and cooperation for radical events; emotions and opinions spilled over virtual communities (crowd sensing), and a holistic view of those individuals who are playing a main role in the event (ideology).

These sets of results can avoid “collateral damages” when they are organised in a “Crisis Scorecard” that works as a cluster of support decision indicators that decide the treatment (course of action) that the specialist (decision makers) will prescribe.

When a disruptive event triggers an online crowd reaction, analysing data from virtual platforms provides a rich source of information for understanding its
Fig. 3: Diagnostic Schema example

genesis and likely evolution. Hence, examining the steps involved using a holistic approach is an essential point for examining Social Movements and detecting Human Security issues.

4 Framework Component Analysis

4.1 Detonating Event

As described in [16], once a political opportunity has triggered a radical societal behaviour, digital information becomes the core asset to identify potential problems.

The preliminary task is to collect those messages or tweets related to the disruptive event; however, as suggested by [16] information flows in four ways: from citizen to citizen, citizen to organisations, organisations to citizens and organisations to organisations. Consequently, selecting the volume of messages that epitomises the disrupting event is the crux of the process.

To tackle this multi-party information exchange process, retweets provide a critical conversational infrastructure as they knit all those voices that need to be heard, and those posts are an adopted practice for those users that want to share and spread thoughts, feelings and ideas to new audiences, as well as trying to engage in conversations [3].

4.2 Warning Period

A central step that needs to be taken relies on detecting the probable danger that follows the detonating incident. As proposed by [20] a disaster can be distinguished according to functional time phases. One of these stages is the Warning Period which refers to the length of time where information reveals a likely menace; however, the detection has to be done just before the aftermath of the crisis becomes perceivable.

In a decision-making scenario, the Warning Period represents a core stage due to the outcome that a correct diagnosis may yield, and can contribute to outlining the “course of action” that has to be followed.
National Security theory comprises a set of complex societal terms, and computational techniques are a valuable tool to solve a high range of problems. Thus in an attempt to couple both concepts to detect potential significant incidents, a two-pronged strategy can be evolved, namely, Event Polarisation and Event Detection.

**Event Polarisation.** As Social Media facilitates the interaction and communication with others [26], people tend to be a primary source of crisis information during a mass emergency event [23], because they use this social software infrastructure to inform their friends, family and acquaintances about their private states (attitudes).

As a result of these set of messages, a significant challenge is to analyse the subjective information to extract and categorise mass opinions that convey a radical idea or oppression feelings [16] and would become the raw material in decision-making.

A computational technique that may be used to detect and analyse Event Polarisation is Sentiment Analysis. This machine learning technique can be used to classify sentiments into three categories: positive, negative and neutral.

The aim to include this process is not limited to detect opinion polarisations, as sentiment fluctuations symbolise the occurrence of sub-events [17], and it can answer questions that surround a collective negative feeling in a selected geographical region.

**Event Detection.** A mass emergency event has a large number of individuals and stakeholders sharing information which is why the volume of messages related to a specific topic increases. However, all these disruptive events are not isolated because they include sub-events [5] that can represent a significant milestone for an effective intervention.

Upon the Event Polarisation process, the system takes each sentiment stream separately (positive, negative and neutral) and extract the topics related to them. A potential clustering method is Latent Dirichlet Allocation (LDA) as it is one of the most popular techniques for this task and has been used to extract topics in major disasters [9].

The next step relies on creating a specialised dictionary that handles words related to human security and is enriched with synonyms to get a reliable wordlist (e.g. ammunition, ammo or munitions).

The fourth step deals with a semantic matching process, where the topics of each cluster is semiantically analysed with the wordlist previously created.

Finally, to identify the nature of the event, this component employs the percentage of topics that are related to each Human Security aspect (economic security, food security, health security, environmental security, personal security, communal security and political security).

A core aspect is that National Security policies are the main reference to evaluate which set of human security components describe a local instability scenario.
4.3 Crisis Interpretation

A common problem that comes after detecting Human Security issues is to create a “big picture” of the disruptive situation. Figure 2 shows in the Crisis Interpretation stage five components that help to interpret radical behavioural elements as well as the way individuals offer or ask for supplies (e.g. money, medicines or weapons) that can be used to help or damage other groups of people.

These components are: Radical Behaviour, Coordination and Cooperation, Communication channels and the Ideology behind opinion leaders.

**Radical Behaviour.** Violence is a radical expression that can be encouraged using social media tools, and during a massive crisis, radical groups tend to distribute their ideology through Twitter users [1].

In accordance with [6], two behavioural markers that describe the way a risk has been increased are: **Leakage** and **Fixation**. The former expresses an intent to harm a specific target (facility, person or any other critical objective).

The second one refers to the tendency to mention with a higher frequency a critical objective; for this paper, Fixation will consider as critical entities: people, facilities, locations and organisations.

Detecting these radical markers is a computational challenge. In our framework, it can be used a natural language processing tool like Named Entity Recognition (NER) to extract the required entities.

Entities are linked to knowledge which is why dissecting the information behind them is a significant aspect. Regarding people, organisations and facilities, fixation can be detected according to the frequency found on tweets.

By contrast, intentions, as explained by [4], are extracted using intention verbs which are associated with an intention action (e.g. “I plan to stay at the Theater”); whereas radical intentions comprise a combination of verbs that keep specific semantic properties. In line with this idea, Levin’s analysis of verbs [12] provide a strong background to create a radical intention structure which is shown in Table 1.

<table>
<thead>
<tr>
<th>Radical Intention Structure</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Levin Verb (Desire)] + [Levin Verb (Killing)] + [Entity]</td>
<td>“I want to eliminate wild animals”</td>
</tr>
<tr>
<td>[Levin Verb (Desire)] + [Levin Verb (Destroy)] + [Entity]</td>
<td>“I desire to burn the Police Station”</td>
</tr>
</tbody>
</table>

As suggested by Levin, verbs of **desire** are: want, crave, desire and need; while verbs of **killing** are: assassinate, eliminate, execute, immolate, kill, liquidate, murder and slaughter and **destroy** verbs are: demolish, destroy, devastate, exterminate and ruin.
Coordination and Cooperation. Microblogging sites offer a broad channel to enhance mass communication during a disruptive event and can be used to express and spread ideas or even radical ideologies and propaganda. However, to achieve shared goals within these virtual communities, collaboration plays a key factor and can be seen as an amalgamation of three main features: communication, coordination and cooperation [7]. Hence, coordination and cooperation create a cooperative system that can monitor the conduct of individuals who interact.

This cooperative system can be divided into two groups: Seekers and Suppliers; both stakeholders are equally important as they can help to detect whether an entity (person or organisation) is reporting their needs or is offering help (e.g. food, medical supplies, water, vehicles, guns, ammunition or money). Natural Language processing and intent mining can deal with both sides of the story as lexical pattern-based structures have been used to solve this issue [15].

Crowd sensing. When mining tweets, people post URL references related to the event they live, and the frequency of these messages suggest the importance of the content. Therefore crawling the information within those websites may discover relevant data (e.g. http://www.libyafeb17.com/?p=916).

The proposed framework requires an iterative loop because the content of these websites needs a complete analysis (Detonating Event, Warning Period and Crisis Interpretation) to identify radical ideology and uncover new events (see Figure 2).

Ideology. Tweets allow individuals and organisations to share not only opinions but pictures, videos, links and email addresses. The later ones represent a source of information to identify likely radical activists. An email address has the following structure: a username and a hostname or domain e.g. username@domain.com (see Table 2).

<table>
<thead>
<tr>
<th>Email</th>
<th>Hostname</th>
<th>DBpedia Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:abc@bbc.co.uk">abc@bbc.co.uk</a></td>
<td>bbc.co.uk</td>
<td>&quot;BBC Online, formerly known as BBCi, is the BBC’s online service&quot;</td>
</tr>
</tbody>
</table>

The username is the crucial element that matches a person against a unique social profile. Twitter and Linkedin are social platforms that have key attributes such as first name, last name, username or date of birth [19]. Assuming that a person X has both online social accounts, a pattern matching process will verify whether these attributes are the same.

Regarding the hostname or domain, it can be matched against large-scale knowledge base such as DBpedia, Wikidata or Freebase to get a “holistic profile” of the person.

Once the profile has been flagged, its content needs a complete analysis (Warning Period and Crisis Interpretation) in order to detect radical behaviour
traits and to understand what kind of coordination and cooperation activity is reigning.

5 Applying Framework on The Libyan Warning Period

Twitter has been a valuable tool used by activists to “overthrow” established governments. Libya made history when Gaddafi’s regime was removed in 2011, and this microblogging service was used to broadcast pictures, telephone numbers, websites and opinions that allowed the escalation of the Libyan uprising.

As a demonstration, this section demonstrates the computational techniques and their results at each sub-stage of the Warning Period: Event Polarisation and the Event Detection (see Figure 2) using as input a set of tweets, related to the Libyan conflict using the hashtag #libya, dated from Feb. 1st to Feb. 28th 2011.

As described in Section 5.1 only retweets were considered, and from a language analysis standpoint, messages written in English were selected. Consequently, the data corpus were reduced from 28,524 to 20,149 tweets.

5.1 Event Polarisation

Before analysing sentiments, the data was cleansed by following three preprocessing steps: 1. URLs, RT and Mention terms were removed; 2. contractions and abbreviations were replaced, and 3. informal ways to convey information (short words) such as: “plz”, “pls”, “ppl”, “peeps”, “pleasert” or “prt” were replaced by its word of origin (e.g. “plz” -> please or “ppl” -> people).

To begin with the Sentiment Analysis process, we used the Stanford CoreNLP library with the Recursive Neural Tensor Network model [18] as our baseline to compute sentiments measures (positive, negative and neutral).

![Fig. 4: Sentiment Orientation](image)

Our data shows that 73% of the analysed messages have a negative polarization (Figure 4), which is why these double figures suggest a clear negative orientation.

Once the polarisation has been detected, the next step focuses on identifying sentiment fluctuations; as described by [17] we calculated the correlation among
(a) Sentiment changes

As can be seen in the Figure 5(a), the ThemeRiver visualization [10] shows that the volume of polarised messages (negative, positive and neutral) increased from February 18th to February 25th, and this gives the chance to identify three visible sentiment shifts (A, B and C) and two essential time frames (Feb. 18th to Feb. 21st and Feb. 21st to Feb. 25th).

Figure 5(b) shows a timeline of the Libyan conflict, outlining the critical events where the sentiment explosions appeared.
5.2 Event Detection

After identifying the sentiment bursts and considering those points as critical subevents, two questions that come to mind are: (1) What topics were conveyed over those time frames? Moreover, (2) are those topics related to National Security?

The first question was tackled by using a topic modelling technique known as LDA which was developed by [2]. However, one of the leading issues lies in determining the number of topics. For this purpose Perplexity analysis was used to evaluate the optimum number of topics, whereas the cross-validation methodology proposed by [8] was used to assess the performance of the topic extraction model.

The second question requires a semantic component to associate those topics that were found in the topic extraction phase, to a Human Security dictionary. One way to deal with this semantic issue was to query the Integrated Public Sector Vocabulary [11], which is a public wordlist that contains a set of terms related to a variety of categories, and some of them are linked to Human Security aspects (Economic Security, Food Security, Health Security, Environmental Security, Personal Security, Communal Security and Political Security). However, to get an enriched dictionary, synonyms were added by using the Wordnet lexical database.

As negative sentiments had a predominant role, the topics extracted from this set of tweets were semantically matched to the expanded dictionary; however, as the volume of tweets had different growth rates and the number of topics was dissimilar overtime, the resulting matched topics were normalised by calculating the percentage of each Human Security aspect per day (see Figure 6).

To understand the way Human Security variables behaved overtime, a Normalised Cross-Correlation analysis was calculated between all of them; but not only in those time frames where the sentiments burst, but a “step before” because it is essential to understand what happened in those previous days. Hence, the Breakout anomaly detection algorithm released by Twitter was used to identify those previous variation points suitably.

As Figure 7 shows, the Breakout algorithm detected two time frames before the changes in sentiment; however, the area shaded in red (AA) and the one shaded in blue (BB and CC) were analysed to understand what happened before and during the sentiment fluctuations appearance.

Table 3 illustrates the cross-correlation results between some of the Human Security aspects such as Public Order, People and Information, and it outlines the following relationships: Public Order – Defence, Public Order – Life, Public Order – Information, People – Defence, People – Health, People – Life, Information – Government and Information – Defence.

According to the Table 3 two scenarios can be identified. The first one (AA to BB) shows that only four out of eight variables had positive increments. On the other hand, the second scenario (BB to CC) shows in as many as seven out of eight analysed variables had positive increases. Hence, the latest scheme suggests that people were strongly engaged in topics related to Public Order and
Table 3: Correlation of Human Security Parameters

<table>
<thead>
<tr>
<th>Topics</th>
<th>Anomaly Detection</th>
<th>Sentiment Fluctuations</th>
<th>Percentage Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feb. 9 - Feb. 16</td>
<td>Feb. 18 - Feb. 21</td>
<td>Feb. 21 - Feb. 25</td>
</tr>
<tr>
<td></td>
<td>(AA)</td>
<td>(BB)</td>
<td>(CC)</td>
</tr>
<tr>
<td>Public Order - Defence</td>
<td>0.2132</td>
<td>0.3133</td>
<td>0.503</td>
</tr>
<tr>
<td>Public Order - Life</td>
<td>0.3109</td>
<td>-0.3919</td>
<td>0.4897</td>
</tr>
<tr>
<td>Public Order - Information</td>
<td>-0.53079</td>
<td>0.08</td>
<td>0.2473</td>
</tr>
<tr>
<td>People - Defence</td>
<td>-0.6181</td>
<td>0.7065</td>
<td>-0.5028</td>
</tr>
<tr>
<td>People - Health</td>
<td>-0.3129</td>
<td>-0.7065</td>
<td>0.6383</td>
</tr>
<tr>
<td>People - Life</td>
<td>-0.1256</td>
<td>-0.2175</td>
<td>0.8116</td>
</tr>
<tr>
<td>Information - Government</td>
<td>0.525</td>
<td>0.3002</td>
<td>0.7492</td>
</tr>
<tr>
<td>Information - Defence</td>
<td>-0.3551</td>
<td>-0.3248</td>
<td>0.8323</td>
</tr>
</tbody>
</table>

Life (224.96%), Public Order and Information (209.13%) and People and Life (473.15%).

As a result, the more positive increments that have been found over a time frame, the more attention that has to be paid to them. This is a key feature that triggers the next stage (Crisis Interpretation). However, the nature of the event and National Security policies will decide which set of Human Security aspects have to be considered to create the Crisis Scorecard and the suitable percentages that have to be reached to activate the next phase.

6 Conclusion and Future Work

This paper has proposed a holistic conceptual framework that utilises computational techniques for examining Social Movements and detecting threats to Human Security.

To demonstrate the feasibility of the framework, the paper has presented a preliminary analysis of tweets related to the Libyan events focussing in particular
on the Warning Period. This analysis has helped to identify two essential frames where critical events occurred. Another core result was the detection of those Human Security variables that had positive variations. This suggests that the more positive increments, the more attention that has to be paid to them because these set of changes showed which aspects epitomised the social disruption.

This preliminary experimental phase has already pointed to some challenges with regard to the components involved in the Warning Period phase. First, slang expressions are still a great challenge because the language has semantic variations from country to country. Hence, creating a robust dictionary may improve radical behaviour detection. Second, spelling mistakes correction is an important issue that NLP has to deal with because it can improve the Event Detection Phase. Third, microblogging platforms tend to spread opinions, but anonymity within this virtual communities is hard to probe.

Going forward, we envisage a robust real-time platform constantly monitoring social media and social behaviour, attempting to identify and predict threats to Human Security. A particular challenge in this endeavour will be the ability to deal with false positives. The next steps are to use different computational techniques and data sets to create and validate an efficient end-to-end monitoring, analysis, predictive and prescriptive responsive platform based on Figure 2. The working system is expected to enable timely and effective responses to possible crises.

References