

Weak Signals as Predictors of Real-World Phenomena in Social Media

Christos Charitonidis, Awais Rashid and Paul J. Taylor
Security Lancaster Research Centre, Infolab21, Lancaster University,
Lancaster, LA1 4WA, United Kingdom
{c.charitonidis, a.rashid, p.j.taylor}@lancaster.ac.uk

Abstract—Global and national events in recent years have shown that online social media can be a force for good (e.g., Arab Spring) and harm (e.g., the London riots). In both of these examples, social media played a key role in group formation and organization, and in the coordination of the group’s subsequent collective actions (i.e., the move from rhetoric to action). Surprisingly, despite its clear importance, little is understood about the factors that lead to this kind of group development and the transition to collective action. This paper focuses on an approach to the analysis of data from social media to detect weak signals, i.e., indicators that initially appear at the fringes, but are, in fact, early indicators of such large-scale real-world phenomena. Our approach is in contrast to existing research which focuses on analysing major themes, i.e., the strong signals, prevalent in a social network at a particular point in time. Analysis of weak signals can provide interesting possibilities for forecasting, with online user-generated content being used to identify and anticipate possible offline future events. We demonstrate our approach through analysis of tweets collected during the London riots in 2011 and use of our weak signals to predict tipping points in that context.

Keywords—Social media, Twitter, Collective action, London riots, Weak signals, Forecasting, Early detection, Content analysis.

I. INTRODUCTION

The tendency of people to form groups and connect with others in order to communicate and engage in collective activities is associated with the human nature [1]. Such groups tend to expand and evolve over time gaining large number of members who share the same purpose or ideas. Social networks are considered the main means of individual recruitment and social movement’s growth [2]. In the age of Internet, and particularly of social media, the collective construction of social movements has become much faster and effortless to achieve [3], [4].

The use of social media as a means of communication and organization of collective action, and particularly as a means of mobilization - in terms of “getting people to the streets”, has become a modern phenomenon. Recent and very relevant examples include London riots and Arab Spring. Social media played a key role, in the former, for arranging times and locations of rioting and looting, while in the latter, in most cases, for organizing peaceful protests.

Detecting the signals of such unfolding phenomena has thus become a key topic of research in the last few years. Recently, several systems have been developed to cope with the problem of event detection in social media. For example, Twitcident [5] analyses information on Twitter for detecting incidents or crises. Similarly, TEDAS [6], a Twitter-based event detection and analysis system, detects new events and

analyses their spatial and temporal patterns. TwitterMonitor [7] automatically detects and analyses emerging trends on Twitter by identifying keywords that suddenly appear in tweets at an unusually high rate. Even though a range of approaches have been proposed, most of them focus on analysing major themes or trending topics at a particular point in time (i.e., strong signals). Consequently, their analysis is undertaken on the “sufficiently visible” topics, leaving only a few or no possibilities for the early detection of emerging events. Others rely on simple counting of keywords/messages, which can provide insufficient insights, or even miss key indicators of such dynamically evolving phenomena.

As yet, little attention has been paid to the identification, detection and analysis of the so-called “weak signals”. Weak signals are early indicators of emerging trends that initially appear at the fringes of mainstream discussion, but quickly and unexpectedly reach a tipping point. Such weak signals have been likened to “hardly discernible cracks anticipating an earthquake” [8]. Weak signals have been proposed as a means of early indicators [8], but no concrete study exists that identifies or validates that such weak signals exist and can be identified reliably.

In this paper, we focus on analysing Twitter data from the London riots in 2011 through a multi-disciplinary approach, drawing upon human behaviour and computer science, to identify and detect weak signals that could be early indicators of the subsequent widespread offline actions. The proposed approach can be of benefit to social scientists and crime analysts in studying online phenomena and their links to offline behaviours.

The rest of this paper is structured as follows. Section II reviews the most related work. Section III presents the proposed model for the detection of weak signals in social media. Section IV gives a brief overview of the 2011 London riots and details the methods used for collecting and analysing the data. Section V presents the results of the data analysis. Section VI gives a discussion of the results and highlights the key characteristics of weak signals. Finally, Section VII concludes the paper and provides directions for future work.

II. RELATED WORK

Most relevant research on mining and analysis in online social media has focused on Social Network Analysis (SNA). SNA is a common approach used to study the relationships and the information exchanged between actors (i.e., individuals or groups). It can reveal groups of actors within a network after examining the relationships existing between them [9]. Identifying how online groups are formed and developed, it is possible to gain predictive insights into future group

behaviour. The patterns of relationships can reveal the way that two or more actors are linked to each other based on their in-between interactions, such as the quantity and frequency of information exchanged. Tie strength allows analysts to examine key groups/actors in the network, as well as to assess the possibility that information will flow from one actor to another. For example, StakeNet [10] uses social network measures to identify and prioritise stakeholders in the network, i.e., individuals or groups that can influence or be influenced.

Rashid et al. [11] analyse social media for detecting masquerading or similarly deceptive behaviour. Using corpus linguistics and natural language analysis techniques, such as Part-Of-Speech (POS) and semantic tagging, the authors predict key characteristics of individuals or groups hiding behind digital personas. POS tagging can provide information about the characteristics of the writing style, while semantic analysis allows classifying words based on their meaning, such as names, locations, emotional states and more. By applying these techniques, it is possible to detect any trends emerging during the conversation, for example, when a conversation is becoming increasingly aggressive.

Another example of what Language Style Matching (LSM) has to offer comes from Taylor et al. [12] who analyse the language changes in emails for detecting insider threats. In a series of simulations, the authors determined that insiders attempting to steal sensitive information did not change their overall work effectiveness (i.e., the strong signal), but they did show subtle changes in “extra-role” behaviours peripheral to their work. These included changes in their self-presentation to others and changes in their emotional engagement with colleagues (e.g., sympathising with work pressures). This leakage may be construed as a weak signal that reflects the psychological changes that occur prior to committing an insider offence against colleagues.

Other recent studies have demonstrated how social media content can be used for predicting real-world phenomena. Bollen et al. [13] analyse the mood of public tweets to predict changes in the stock market. Their results showed that an emotional change in tweets is followed by a similar rise or fall in the stock market prices from 2 to 6 days later. This indicates that mood (sentiment) changes in social media can possibly have predictive value regarding real-world phenomena. Sakaki et al. [14] propose an approach for detecting earthquakes in Twitter based on a probabilistic spatio-temporal model. The authors achieved to send email notifications about earthquakes much faster than the meteorological agencies. In [15] authors show how time-series and sentiment analysis on tweets related to movies can actually predict box-office revenues. [16] presents a framework, namely SNEFT, which predicts real-world flu epidemics through social media analysis. The authors show that the number of flu-related tweets is highly correlated with the rate of influenza cases reported by official statistics.

In contrast to the above studies, our work aims at using publicly available user-generated content to detect weak signals that can be early indicators of real-world collective phenomena, such as civil unrest. Additionally, as opposed to other works, our work is not limited to a single dimension in analysing the data, but considers a multi-dimensional approach consisting of various data analysis methods - both quantitative and qualitative.

III. PROPOSED PREDICTIVE MODEL

The proposed model (Fig. 1) consists of a variety of techniques for the detection of weak signals in social media. These techniques include (i) keyword analysis, (ii) geo-spatial analysis, (iii) frequency analysis, (iv) semantic analysis and (v) sentiment analysis. Using these techniques we aim to detect weak signals and their patterns that could lead to the prediction of future offline phenomena. For example, the detection of geo-location information (either in the form of coordinates or location names in text) in a higher frequency than in the past could be a form of a weak signal. Also, keyword and semantic analyses play a key role when it comes to detecting words and topics of interest. Semantic analysis allows the detection of further items that would be harder, if not impossible, to spot at the keywords level in such large-scale datasets. A word, or semantic domain, is *key* if it occurs more often than would be expected by chance in a corpus compared to some norm [17]. Key words/semantics that initially appear at the fringes of online communications, but quickly gain strength over time, can be early indicators of future trends. Finally, sentiment analysis can significantly improve the prediction of events by giving insights into the emotions of users and help in detecting emotional shifts. The negative sentiment is usually the key to widespread events in Twitter; as such events are associated with rises in the negative sentiment [18]. Therefore, unusual rises in the negative sentiment could be an early indication of emerging trends.

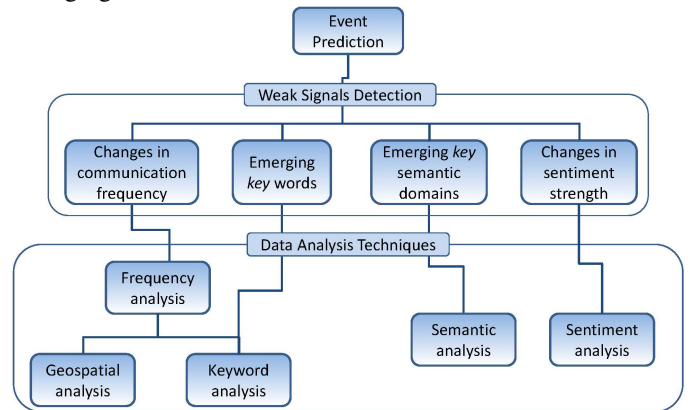


Fig. 1: Proposed predictive model

IV. EVALUATION METHODOLOGY

We evaluate our proposed model through analysis of tweets collected during the London riots in 2011.

A. The 2011 London Riots

Between 6 and 10 August 2011 riots took place across England in which buildings were burned down, shops looted, people were killed and property damages amounted to hundreds of millions of pounds. A suspect’s shooting¹ by police on 4 August at around 18:15, had sparked the riots which initially started as a peaceful protest in Tottenham two days later (6 August), but within hours it turned into massive violence and opportunistic looting². Soon after, the riots spread throughout London with some of the most affected areas being

¹<http://www.theguardian.com/world/2011/aug/05/man-shot-police-london-arrest>

²<http://www.theguardian.com/uk/2011/aug/07/tottenham-riots-peaceful-protest>

Woodgreen, Enfield and Walthamstow. The use of social media not only helped rioters in organising these events, but also caused more people to find out about the riots and get involved, as well as inspired more violence in other locations too - what described by the media as *copycat violence*.

B. Data Collection

Using Topsy³ we collected publicly available tweets posted between 4 and 10 August 2011 (totally, 831,041 tweets posted by 361,200 unique users). The rationale behind collecting tweets - if any - two days before the riots started is to observe any potential changes in the emotions, behaviour, communication frequency, etc.; particularly, after the shooting occurred. The most popular and trending topics during the riots were selected, i.e., #tottenham and #londonriots. We queried Topsy's database with combinations of keywords like: (i) #tottenham OR tottenham and (ii) #londonriots OR (#london AND riots), and used language-filtering to exclude any non-English tweets. The restriction to English was selected to avoid any complications of multiple languages in the results. Once tweets were downloaded, we extracted additional information from them, such as user mentions, hashtags, urls, timestamp (converted to GMT+1), etc. From the collected tweets, we formed two datasets, i.e., (i) *Tottenham* and (ii) *London riots*.

C. Data Analysis

First, we use frequency analysis to identify the times at which communication frequency changes and detect any potential correlations between tweets' rate and the actual offline events. Further, we use geo-spatial analysis by extracting any location information embedded in tweets (i.e., words referring to locations) and use the Google charts API to generate heat maps of the initially online mentioned, but later actually affected locations. We also use SentiStrength⁴, a sentiment analysis tool, to detect any potential emotional changes in the communication before, during and after the incidents. SentiStrength is designed especially for estimating the sentiment in *short* texts (even for informal language) and can also deal with misspellings, emoticons, etc. It is thus a suitable tool for analysing the sentiment of *tweets*. SentiStrength classifies the text simultaneously as positive and negative by assigning a value for each text on a scale from +/-1 (*no* sentiment) to +/-5 (*very strong* positive/negative sentiment) [18]. Due to the high volume of tweets, we first classified all tweets by SentiStrength and then calculated the average positive and negative sentiment scores of all tweets per time-scale (e.g., per quarter-hour). We extend the analysis further by using natural language analysis techniques to detect *key* words and topics/themes within the conversations. This analysis is aided by Wmatrix⁵, a corpus analysis and comparison tool. Wmatrix has been applied to numerous studies including political science research, online language analysis, corpus stylistics, as well as extremist language analysis [11], [17], [19]. Amongst other things, Wmatrix allows comparing a corpus (i.e., a set of texts) with a reference corpus (or another subset of the dataset) to obtain 'key' items at different domain levels, such as *word* and *semantic*. In this paper, the *BNC Sampler Corpus Written*⁶ was used as our reference

corpus. The *BNC Sampler Corpus Written* is a sub-corpus of the British National Corpus, consisting of approximately 1 million words from a wide variety of written British English. It contains a wide and balanced sampling of texts from the BNC, maintaining the general text types (and their proportions) of the BNC as a whole. It is thus suited for comparing the collected English tweets with it. Wmatrix compares the relative frequencies of occurrence of items in the two corpora using the log-likelihood (LL) statistic [17]. Items with a LL value of 6.63 or higher (i.e., at the significance level of $p < 0.01$) are considered to be 'key' items. The most significant 100 key items are visualised as "tag cloud" in alphabetical order. The larger an item is, the greater its significance. Wmatrix also allows qualitative examination of the key items through the use of concordances by clicking on items within the cloud. This allows us to closely examine the context behind the key items and get more qualitative insights into users' thoughts, feelings and behaviour. Before undertaking the analysis with Wmatrix, tweets were converted to lowercase and cleansed of all information that would lead to mistaken results. More precisely, the cleaning process involved the removal of special characters, prefix "RT" (i.e., abbreviation of the word "re-tweet"), usernames and urls.

V. RESULTS

A. The Tottenham Riots

Fig. 2(a) shows the hourly number of tweets from *Tottenham* dataset between 4 and 10 August 2011. As can be seen in Fig. 2(a), while the flow of tweets was initially stable and low in frequency, however, in the evening of 4 August it was followed by a slight increase. Fig. 3(a), which depicts the hourly average sentiment of all tweets from *Tottenham* dataset, shows that there has been a sharp increase in the negative sentiment during that time-period, whereas, at the same time the positive sentiment fell (-ve=-2.18 and +ve=1.22 on 5 August at 01:00). The linguistic analysis of tweets, from 18:00 on 4 August to 01:00 on 5 August, shows that this weak, however, important rise on tweets' rate is strongly associated with the incident that occurred. Fig. 4(a) illustrates the tag cloud with the most significant keywords as calculated by Wmatrix. Keywords, such as *shooting* (LL=743.38), *dead* (LL=549.22), *police_officer* (LL=450.59) and *police* (LL=389.21) are among the top overused ones. Semantic analysis (Fig. 5(a)) also shows the categories *Warfare_defence_and_the_army;* *weapons* (LL=846.32), *Dead* (LL=381.59), *Law_and_order* (LL=317.83), *Time:_Present;* *simultaneous* (LL=20.13) and *Vehicles_and_transport_on_land* (LL=64.73) to be among the predominant key semantic domains. A concordance analysis of the content behind these key items (both words and semantics) shows clearly that the conversation revolves around the *incident* (LL=142.98) and the *man* (LL=547.80) who *got* (LL=73.96) *shot* (LL=779.70) at *Tottenham hale* (LL=760.98). Specifically, one user posted - "Arrrggggh The guy who got shot in Tottenham Hale died :(" (4 Aug 20:13). "Another death in tottenham! Smh!!" (4 Aug 22:57), wrote another. The emotional shift, along with the *key* words and themes, indicate weak signals of the growing collective anger towards the police; which would eventually become one of the major reasons for the riots⁷. These early signals could have given insights into the behaviour and emotions of people

³<http://topsy.com/>

⁴<http://sentistrength.wlv.ac.uk/>

⁵<http://ucrel.lancs.ac.uk/wmatrix/>

⁶<http://ucrel.lancs.ac.uk/bnc2sampler/sampler.htm>

⁷<http://www.theguardian.com/uk/2011/dec/05/anger-police-fuelled-riots-study>

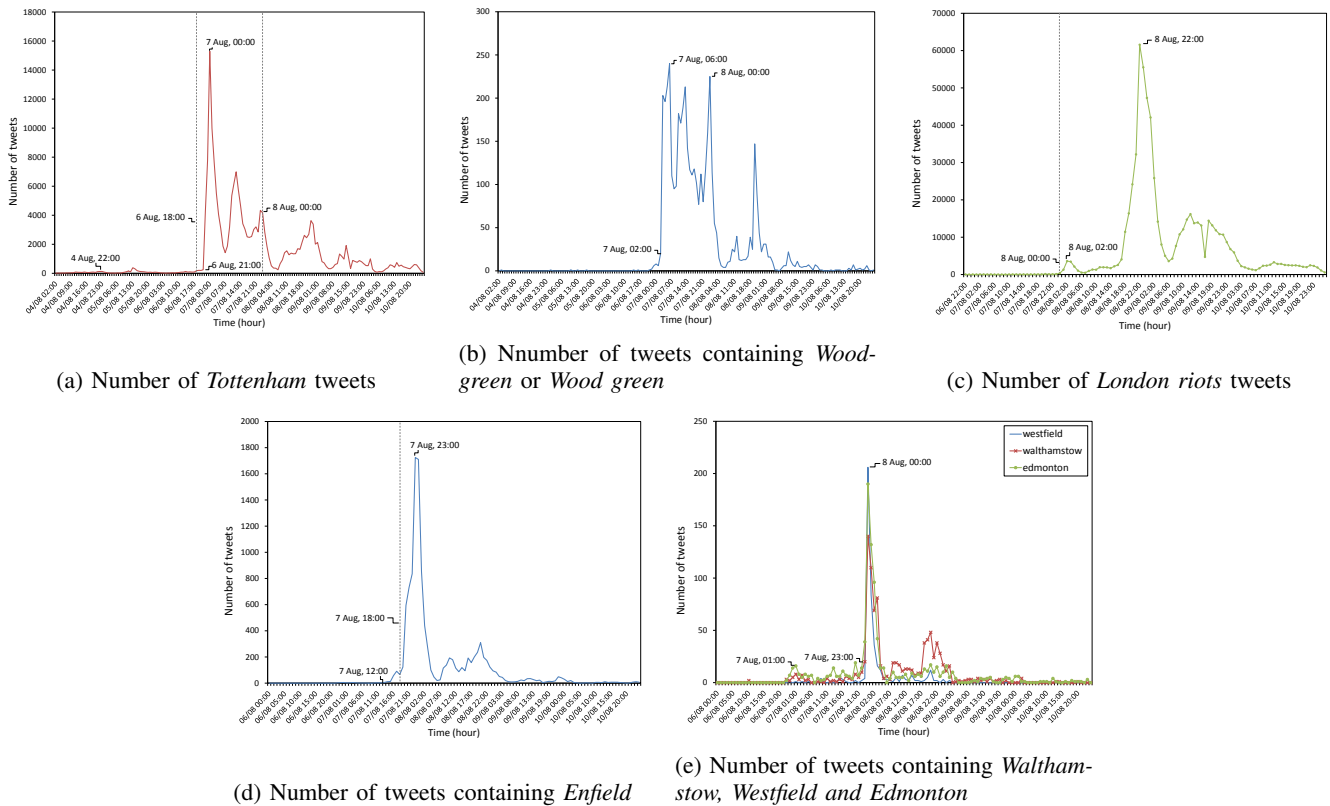


Fig. 2: Frequency analysis

before the protest started on 6 August - which would shortly turn violent.

On 6 August, as shown from Fig. 2(a), there has been a gradual increase from 18:00 to 20:00, while after 21:00 tweets' rate dramatically rose. Figures 4(b) and 5(b) show the most significant key words and semantics between 18:00 and 20:00. At first glance, it seems that the whole discussion revolves around sports. However, examining the key items more closely, it can be seen that the discussion about the protest appears at the fringes of this mainstream topic. Content analysis of keywords shows that tweets refer to the *march* (LL=47.15) that is *kicking_off* (LL=42.49) *right_now* (LL=52.51) outside the *police_station* (LL=54.75) on Tottenham *high* (LL=99.99) *road* (LL=92.80). Key semantics (appearing at the periphery) include *Time:_Present;_simultaneous* (LL=21.75), *Time:_Beginning* (LL=17.62), *Law_and_order* (LL=7.88), *Vehicles_and_transport_on_land* (LL=29.20) and *Violent/Angry* (LL=44.88) (Fig. 5(b)). All these semantics indicate the beginning of the protest, as well as the likely ensuing violence. Tweets - many of them of aggressive and hostile nature - encourage others to join in. For instance, one user posted: "If you hate police come Tottenham high road now" (6 Aug 19:05). These early signals show that the discussion around the protest was becoming an emerging topic. Also, they give insights regarding the hostility of some people towards the police and their efforts to mobilise the public.

As shown in Fig. 3(a), after 20:00 on 6 August, the negative sentiment sharply increased while the positive sentiment slightly dropped. Analysing the tweets from 20:00 to 21:00 on 6 August (Fig. 4(c)), it appears that keywords such as *police_station* (LL=83.65), *right_now* (LL=101.29) and *riot* (LL=281.43) appear significantly

stronger than in the previous hours. Tweets refer to the group of people who attacked and set on *fire* (LL=58.79) two *police* (LL=143.48) *cars* (LL=68.22) (i.e., the time that the peaceful protest turned violent). Key semantic tags (Fig. 5(c)) include *Law_and_order* (LL=56.64), *Time:_Present;_simultaneous* (LL=11.11), *Temperature:_Hot/_on_fire* (LL=68.21), *Vehicles_and_transport_on_land* (LL=66.15), *Warfare,_defence_and_the_army;_weapons* (LL=9.72) and *Violent/Angry* (LL=107.87). All these semantic domains highlight the situation predominated that time outside the Tottenham's police station. Although the discussion about the riots was still at the fringes, it was increasingly emerging and would shortly pass to the mainstream.

As can be seen from Fig. 2(a), the volume of tweets rapidly climbed after 21:00 on 6 August and reached a peak by the end of the day. As Fig. 4(d) (21:00 - 22:00 on 6 August) shows, the conversation about the *disturbances* (LL=1061.44) outside the *police_station* (LL=557.28) had moved from the fringes and became the trending topic on Twitter. Keywords such as *police* (LL=6350.46), *riots* (LL=3209.72), *shooting* (LL=1023.18) and *right_now* (LL=963.56) are among the top significant key words. From the semantic analysis on this time-period (Fig. 5(d)), it can be clearly seen that the main theme is the on-going riots. Key semantic domains include *Violent/Angry* (LL=6194.84), *Warfare,_defence_and_the_army;_weapons* (LL=443.39), *Temperature:_Hot/_on_fire* (LL=2903.46), *Crime* (LL=148.05), *Law_and_order* (LL=3936.67), *Vehicles_and_transport_on_land* (LL=3175.43), *Time:_Beginning* (LL=122.79) and *Time:_Present;_simultaneous* (LL=698.89). The content analysis shows that people not only distributed the news, but they also incited the crowd to violence and looting. Examples

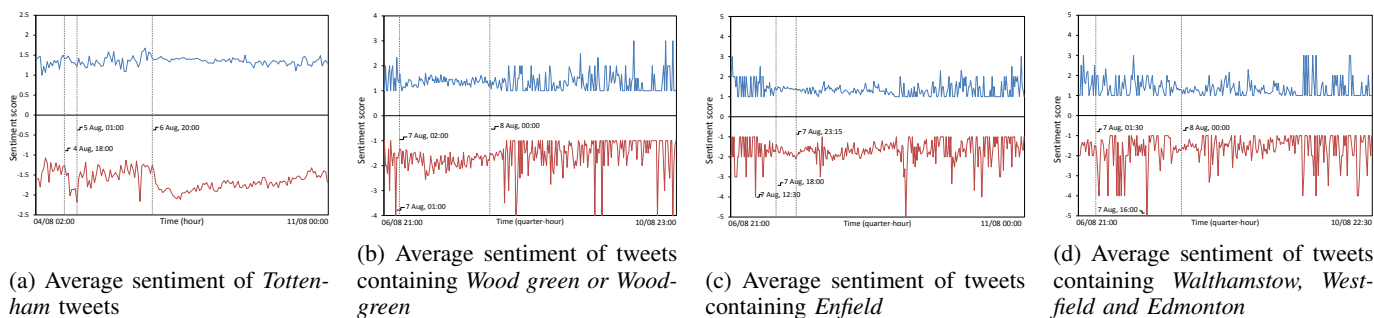


Fig. 3: Sentiment analysis

include: (i) 6 Aug 20:58 - “Who wants to come loot with me in tottenham?” (ii) 6 Aug 22:24 - “En route to Tottenham to break into the jewellery shop”. These signals could have given insights into the escalation of the situation that would follow in the next hours, as well as possible locations of future attacks by rioters.

In the early hours of 7 August, the nearby Woodgreen was affected by riots; whereas police was still not present at 04:00⁸. As can be seen in Fig. 2(b), from 20:00 on 6 August to 02:00 on 7 August, there has been an unusual growth in the number of tweets. The sentiment analysis on this time-period (Fig. 3(b)) shows a steep rise in the negative sentiment, reaching a score of -4 on 7 August at 01:00. These spikes in the negative sentiment indicate a possible emerging event. Before uploading the texts to Wmatrix, occurrences of “Wood green” were converted to “Woodgreen” to avoid any mistaken results in the analysis. Analysing the tweets from 20:00 on 6 August to 02:00 on 7 August (Fig. 4(e)), it appears that users warn that rioters (LL = 14.45) have moved (LL = 22.44) from Tottenham and they are tryna (LL = 28.90) start (LL = 17.14) a riot (LL = 102.53) in Woodgreen N22 (LL = 28.90). Moreover, they mention incidents they eye-witness, such as that people smash_up (LL = 14.45) the t-mobile (LL = 14.45) shop. Semantic analysis (Fig. 5(e)) shows categories such as Location_and_direction (LL = 7.97), Crime (LL = 6.68), Time:_Future (LL = 7.86), Time:_Present;_simultaneous (LL = 6.78) and Violent/Angry (LL = 48.04) being among the emerging themes.

These emerging key items (words and semantics) give early warnings of the impending riots in Woodgreen. After 02:00 on 7 August, tweets’ rate sharply increased and reached a peak by 06:00 (Fig. 2(b)). As can be seen in Fig. 3(b), the negative sentiment consistently outweighed the positive sentiment in the following hours, giving insights into the current emotions of users. Keyword analysis between 05:00 and 06:00 on 7 August (Fig. 4(f)) shows that riots (LL = 541.06) had moved to Woodgreen and become a trending topic. Specifically, people report that the shopping_centre (LL = 57.69) is being looted (LL = 188.91) right_now (LL = 165.72) by youths (LL = 152.07) wearing masks (LL = 125.94). The semantic analysis (Fig. 5(f)) shows the emergence of the previously peripheral key domains such as Crime (LL = 212.42), Time:_Present;_simultaneous (LL = 66.69) and Violent/Angry (LL = 363.15). In addition, among the top key concepts are Law_and_order (LL = 120.43), Time:_Beginning (LL = 33.04), Temperature:_Hot /_on_fire (LL = 62.46), Damaging_and_destroying (LL = 47.26) and Ve-

hicles_and_transport_on_land (LL = 13.15).

B. The London Riots

While the riots in Tottenham were fading out, disturbances would start spreading across other areas of London (Fig. 7(b)); with the situation being escalated in the evening of 8 August (Fig. 7(e)). As can be seen from Fig. 2(a), the volume of Tottenham tweets quickly went down after 00:00 on 8 August. However, keyword analysis of Tottenham tweets after 00:00 on 8 August reveals the gradual rise of the keyword londonriots at the times: 00:00 (LL = 140.09), 01:00 (LL = 368.76) and 02:00 (LL = 1263.35). This early indication could have given warnings for the following shift from Tottenham to London riots and the impending massive violence throughout London. As can be seen in Fig. 2(c), there has been a sudden rise in the number of tweets (from London riots dataset) between 00:00 and 06:00 on 8 August. The londonriots keyword was accompanied by names of London boroughs that were about to get affected by riots in the following hours.

Some of these areas were Enfield, Walthamstow, Westfield and Edmonton. Fig. 2(d) shows the hourly frequency of tweets containing Enfield. As shown in the figure, between 12:00 and 18:00 on 7 August, the number of tweets suddenly went up. That rise (as can be seen in Fig. 3(c)) is connected to a rise in the negative sentiment too. Keyword analysis in Fig. 6(a) shows people on twitter spreading rumours (LL = 162.46) saying that riots (LL = 588.21) are due to start (LL = 68.59) in Enfield at 4pm (LL = 190.18). Others hear that more trouble (LL = 53.69) is expected in Enfield tonight (LL = 61.68). Key semantic categories (Fig. 6(d)) include Crime (LL = 45.80), Damaging_and_destroying (LL = 31.02), Law_and_order (LL = 111.23) Time:_Present;_simultaneous (LL = 118.08) and Violent/Angry (LL = 484.65). As can be seen from Fig. 2(d), the volume of tweets considerably rose after 19:00 on 7 August and reached a peak by 23:00. Keyword analysis, from 22:00 to 23:00 on 7 August (Fig. 6(b)), shows that riots had moved to Enfield. People report that protestors (LL = 914.32) are throwing (LL = 1057.91) petrol_bombs (LL = 1098.93) on passing (LL = 911.51) cars (LL = 905.48). Others describe incidents unfolding next to them, such as that looters (LL = 173.62) are looting (LL = 732.34) the tesco (LL = 72.34) and krispy kreme (LL = 376.18) shops. Semantic analysis (Fig. 6(e)) shows domains such as Violent/Angry (LL = 1533.72), Time:_Present;_simultaneous (LL = 264.69), Crime (LL = 1588.14), Vehicles_and_transport_on_land (LL = 458.30), Damaging_and_destroying (LL = 86.54) and Warfare_defence_and_the_army;_weapons (LL = 252.86) being the prevalent themes.

Fig. 2(e) shows the rate of tweets containing Walthamstow,

⁸<http://www.theguardian.com/uk/2011/aug/07/tottenham-riot-looting-north-london>

arsenal baffled...bbc...blackburn...brad...chelsea...christian...city...comcoys...dead...diarra...erik...europa...fernando...ferry...friedel...goalkeeper...guy...hale...happened...haringey...hotspur...huff...incident...injured...ipcc...joey...lane...liverpool...lorente...london...ltmadrid...man...modric...n...north...officer...police...police...officer...shooting...shots...signing...stoke...thfc...thorsted...tottenham

(a) Tottenham (4 Aug 18:00 - 5 Aug 01:00)

0_1_25_2_1...50...53...athletic...athl...atic...balo...bilbao...bolbao...city...friendly...fr...crouch...defoe...diarra...everton...football...game...garry...glory...goals...hater...hearing...high...ho...tspur...hotspurs...ht...jermain...jose...love...lt...madrid...march...modric...peter...riof...road...san...season...spurs...stake...stand...starting...thfc...tottenham...white...watching

(b) Tottenham (6 Aug 18:00 - 20:00)

122_1...athletic...bilbao...crouch...dandefoe...der...dugan...enough...favour...feds...fire...fuck...lacking...given...guardian...hale...happening...harry...hero...high...hotspur...jermain...k...lub...liverpool...london...luka...madness...man...mo...dric...peter...pinaar...pinaar...injury...police...police...station...redknapp...rescue...right...now...riot...riots...spanyol...spurs...steven...string...t...ekuk...tottenham...welcome

(c) Tottenham (6 Aug 20:00 - 21:00)

attacked...bbc...burning...car...cars...crazy...disturbances...duggan...during...fire...happening...in...kicking...off...london...ltmad...madness...man...news...north...on...over...shooting...people...pics...police...p...olice...station...protest...right...now...riot...rioting...riots...road...set...o...n...fire...shit...shooting...shot...sky...tottenham

(d) Tottenham (6 Aug 21:00 - 22:00)

4_next...boys...bredrins...hearing...i...just...live...lol...moved...n22...na...no...way...riot...riots...roiting...s...safes...stow...t...ham...tottenham...tryna...used...woodgre...en

(e) Woodgreen (6 Aug 20:00 - 7 Aug 02:00)

bbc...brainwashing...bristol...damage...differently...doing...done...events...fishy...gleeful...hashtags...igniting...lpc...lawlessness...lbc...london...looted...looters...looting...masks...news...on...site...over...pierce...police...pounds...port...reporting...right...now...riot...rioting...riots...shopping...centre...sirens...surprise...tottenham...tottenhamriot...twitter...tw...itter...youths

(f) Woodgreen (7 Aug 05:00 - 7 Aug 06:00)

Fig. 4: Key word clouds

Dead...Disease...Evaluation...Bad...Geographical...n...ames...Law...and...order...People...Male...Sports...Unmatched...Unselfis...h...vehicles...and...transport...on...land...Warfare...defence...and...the...army...weapons

(a) Tottenham (4 Aug 18:00 - 5 Aug 01:00)

Children's...games...and...toys...General...appearance...and...physical...properties...Geographical...n...ames...Other...proper...names...People...Male...Sports...Sports...and...games...Unm...atched...Unselfish...Vehicles...and...transport...on...land...Violent...Angry

(b) Tottenham (6 Aug 18:00 - 20:00)

Geographica...I...names...Interested...excited...energetic...Law...and...order...Long...tall...and...wide...Sports...Success...Temperature...Hot...on...f...ire...Unmatched...Vehicles...and...transport...on...land...Violent...Angry

(c) Tottenham (6 Aug 20:00 - 21:00)

Alive...Arts...and...crafts...Change...Crime...Discourse...Bin...Disease...Evaluation...Geographical...names...Happy...Judgement...of...appearance...Ugly...Knowledge...eable...Law...and...order...Male...Temperature...Hot...on...fire...Unmatched...Vehicles...and...transport...on...land...Violent...Angry...Warfare...defence...and...the...army...weapons

(d) Tottenham (6 Aug 21:00 - 22:00)

Geographical...names...Happy...Location...and...direction...Negative...Residence...safe...Time...Unmatche...d...Vehicles...and...transport...on...land...Violent...Angry

(e) Woodgreen (6 Aug 20:00 - 7 Aug 02:00)

Alive...Business...Selling...Clothes...and...personal...belongings...Cr...ime...damaging...and...destroying...Geographical...names...Law...and...order...Temperature...Hot...on...fire...Unmatched...Vehicles...and...transport...on...land...Violent...Angry

(f) Woodgreen (7 Aug 05:00 - 7 Aug 06:00)

Fig. 5: Key semantic clouds

Westfield and Edmonton. As shown in the figure, in the early morning of 7 August, these areas suddenly started emerging in the conversation on Twitter; whereas, after 23:00, the volume of tweets rapidly increased. The sentiment analysis (Fig. 3(d)) shows a similar pattern (i.e., sudden spikes in the negative sentiment) to the previously analysed boroughs before the event escalated. As can be seen in Fig. 3(d), the negative sentiment sharply increased after 01:30 on 7 August, reaching high negative scores. Keyword analysis (from 21:00 on 6 August to 23:00 on 7 August) shows that people are hearing (LL=41.19) that Westfield is going to be hit by riots. One wrote, "... these broadcasts telling me to meet up to trash westfield" (7 Aug 14:46). Others report that they are getting (LL=40.28) messages saying that riots are going to kick off in Waltham-

stow. In Edmonton, people report that the jib (LL=30.78) shop is being looted (LL=61.57) while police (LL=78.74) is dealing with riots in Tottenham. Shortly after 22:00 on 7 August, people report that a massive (LL=67.43) group of youths (LL=119.29) is gathering (LL=78.40) in Walthamstow beginning (LL=42.75) to smash up (LL=112.87) a shop (LL=94.51). Semantic analysis Fig. 6(e) shows categories such as Crime (LL=98.47), Damaging and destroying (LL=38.03), Vehicles and transport on land (LL=29.45), Time: Present; simultaneous (LL=113.09), Time: Beginning (LL=14.26) and Violent/Angry (LL=335.16) being amongst the top key topics. These weak signals could clearly provide early warnings about the beginning and the impending escalation of riots throughout these areas in the following hours.

4pm amid anarchy apparently being breaking City closed coming to
 damaged due early edmonton enfield evacuated
 evacuating Everyone leaving fro go gon
 lo! london looting .madness mandem mcdonalds na north
 hing. to. do. with now in people police protests reports ridiculous riot
 rioters rioting riots rumour r umour r umours safe scam s
 hit shops shutting spread spreading spreads start in the shops. tonig
 ht tottenham town trending trouble tweets twitter
 u walthamstow westfield woodgreen

(a) Enfield (7 Aug 12:00 - 18:00)

a10 alert are avoid brixton cameron cars edm
 onton enfield
 st kicking off kreme krispy last night later live lol london look at l
 oot looter looters looting it may n't nandos news north nownt olympics omg
 on passing people petrol_bombs police posing prote
 sters protestors regret riot riot rioters rioting riots roa
 d scum SHIT sirens skynews via stupid tweets throwing thugs tonight tottenham
 ule walt walt twitter u uk updates v via violence waltham war. zanzibar why woodgreen
 wtf

(b) Enfield (7 Aug 22:00 - 23:00)

Business: Selling Closed_HidingHidden Conspiracy Crime Damaging_and_de
 stroying Degree Evg Evaluation: The Excluzivizers/particularizers Expected Foolish G
 eographical_names Law_and_order Location_and
 e: Present; simultaneous Time: Early Time: Momentary
 Beginning Unconnected Unmatched Violent/Angry

(d) Enfield (7 Aug 12:00 - 18:00)

Alive Attentive Avoiding Comparing_SimilarDifferent Crime Damaging_and_destr
 oying
 Degree Discourse_Bin Distance Evaluation: Bad Excluzivizers/particularizers
 Expected Failure Foolish General_appearance_and_physical_propertie
 s Geographical_names Happy Inability/unintelligence Judgement_of_
 appearance: Ugly Knowledgeable Law_and_order Lawful Location_and_directio
 n Money_and_pay Negative No_respect No_obligation_or_necessity No_contentment Participating People Personal
 ty_traits Putting_pulling_pushing_transporting Relationship: Intimacy_and_sex Relationship: Sexual/Reside
 nce Safe Sad Seen Sensory_Sight Sensory_Sound Temperature: Hot / on fire The Media_TV_Radio_and_Cinema The universe Tim
 e: Late Time: Present; simultaneous Time: Momentary
 Time: Past Time: General Unethical Unmatched Vehicles_an
 d_transport_on_land Violent/Angry Warfare_defence
 and_the_army;_weapons Worry

(e) Enfield (7 Aug 22:00 - 23:00)

m 2 ambulances anyone apparently as well beginning brixton chingfor
 d cking confirm cos cryed edmonton enfield exclusive from fuckers gathering g
 et getting gon gon gt hackney hearing honey i iceland ikea im images in ju
 st kicking off kick down system live lol london lool lool looted loot
 ng lt massive my n n n next north now nt partners peckham playstation police
 ponds_and ppl praying riot rioters rioting riots shop sirens smash_u
 p spread square suspect tomorrow tonight tottenham totty trouble
 u walthamstow westfield woodgreen y
 ouths

(c) Walthamstow, Westfield, Edmonton (6 Aug 21:00 - 7 Aug 23:00)

Alive Business_Selling Comparing_SimilarDifferent Crime Damaging_and_destr
 oying
 Discourse_Bin Distance_Nearness Evaluation: Bad Excluzivizers/particularizers Ex
 pected Failure Foolish Geographical_names Happy Lawful Ligh
 t Location_and_direction Negative Resi
 dence Safe Sensory_Sound Temperature: Hot / on fire Time: Presen
 t; simultaneous Time: Beginning Time: Unmatched Ve
 hicles_and_transport_on_land Violent/Angry

(f) Walthamstow, Westfield, Edmonton (6 Aug 21:00 - 7 Aug 23:00)

Fig. 6: Key word and semantic clouds

VI. DISCUSSION

Our results demonstrate that weak signals indeed exist in social media, such as Twitter, and that the proposed model can be successfully used in the detection of these early indicators for predicting future trends and phenomena.

From our analysis, we have derived predictive patterns, such as changes in communication frequency, emotion and language, that could constitute the characteristics of weak signals in social media. For instance, the rate of messages on Twitter seem to follow a path that is closely related to emerging events in the real-world. Thus, when the rate of tweets turns unusually high compared to the normal rate, it could be an indicator of a weak signal. Our findings also show that such a sudden rise in the volume of messages is usually accompanied by spikes in the negative sentiment - which could be another key characteristic of weak signals. This correlation between rises in frequency and sentiment could form a pattern of weak signals.

It's noticeable how Twitter users instantly report what's happening next to them by sharing images/videos and being specific by giving locations. It is also interesting the way they share broadcast messages from other social networking services in an effort to warn the general public for upcoming disorders in specific locations. Thus, a geo-spatial analysis in the case of such phenomena proves to be valuable. Performing frequency analysis over geo-spatial analysis, we can detect early indications of emerging events in different geographical areas.

The word and semantic analysis has also provided us with key words and themes that we could look out for in weak signals of interest (i.e., in the context of civil unrest). For instance, *Violent/Angry* domain seems to be one of the emerging key themes in such phenomena that increasingly becomes stronger over time. This semantic domain is associated with words and phrases that indicate violent actions and aggressive behaviours, such as *riot*, *disturbances*, *fight*, *kick*, *attack*, etc. If this domain, or words, are emerging within the online conversations, then they could be early indicators of impending

disturbances and violent actions. Similarly, the *Crime* domain, which is associated with criminal behaviour such as *looting*, *stealing*, *breaking into*, etc., as well as the *Damaging_and_destroying* domain which is connected to *breaking*, *smashing_up*, *destroying*, etc. In addition, *Law_and_order* was one of the predominant key themes, since *police* is usually directly related to such events. Thus, the emergence of this theme can give early indications of possible aggressive behaviours against the police. *Geographical_names* is also one of the key domains that could be used in conjunction with geo-spatial analysis; since this semantic domain includes information related to locations and places. The detection of *key* geographical names can provide early indications of emerging local events. Moreover, the *Vehicles_and_transport_on_land* domain, which includes terms such as *vehicles*, *cars*, *streets* and *road*, seems to be a key topic in such phenomena. Last, but not least, *Time: Present; simultaneous* and *Time: Beginning* domains can provide early warnings about events happening *right now*, *today*, *tonight*, or events that are *starting*, *kicking off*, *beginning*, etc.

Paying attention to these key characteristics of weak signals can improve the detection of signals of interest in large-scale social media data. The proposed model could possibly be applied in detecting weak signals in domains other than civil unrest; since the key word and semantic analyses are not domain-restricted.

One of the limitations of this study was the lack of coordinates in our datasets from any possible geo-tagged tweets. Although the amount of authors having enabled the geo-tagging feature would normally be small, we could possibly get more precise information as to where some of these tweets were coming from while the riots were forming. Also, if the collection of tweets containing the affected areas was performed as a separate query on Topsy, the detection of weak signals could significantly be improved - since in this analysis the names of these areas had to co-occur with the words we queried the Topsy database.

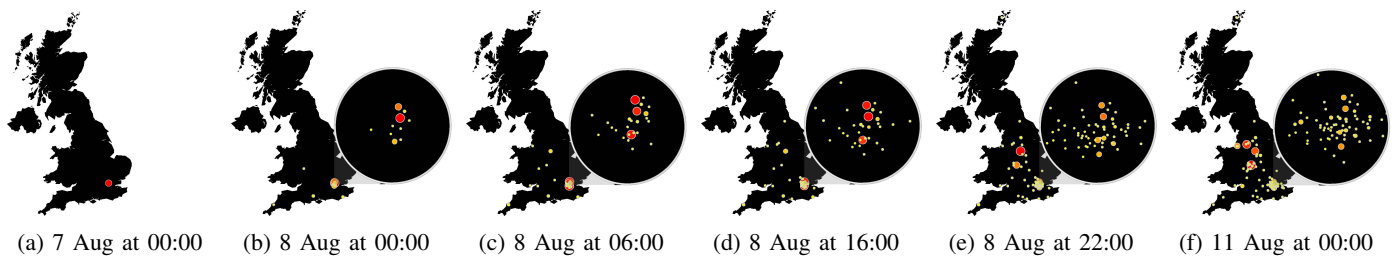


Fig. 7: Locations mentioned in tweets and which actually affected by riots

VII. CONCLUSION AND FUTURE WORK

Undoubtedly, social media facilitate the organisation of events and interaction among masses of people in a way that no other communication tool did before. They increase the possibility for leaders to exercise their influence on others, as well as give voice to introverts, i.e., people that would normally stay silent. More importantly, Twitter seems to act as a valuable reporting tool for anyone who can instantly report and give information from their position.

In this paper, we analysed Twitter data from the London riots in 2011 using a multi-disciplinary approach in an attempt to identify and detect weak signals to predict tipping points in that context. The results validate that weak signals indeed exist in social media and can be used for predicting real-world phenomena, such as the London riots. The detection of such weak signals in social media, could inform law enforcement agencies to take precautions, as well as countermeasures, in the case of impending violent collective actions.

In future, we aim to examine the usefulness of social network analysis techniques in identifying potentially predictive behaviours hidden in the connections between individuals and groups. We will examine the importance of these connections and actors (both central and peripheral) in spreading the information in the network and influencing their neighbours. Finally, we aim to use collocation analysis techniques (provided by Wmatrix, at the word and word-semantic levels) to examine networks and links between words, actors and geographical places, etc., to identify possible weak signals and their patterns for forecasting future events.

ACKNOWLEDGEMENTS

We would like to thank Paul Rayson for giving us access to Wmatrix and Mike Thelwall for providing us with the SentiStrength Java version.

REFERENCES

- [1] D. R. Forsyth, *Group Dynamics*. Wadsworth/Cengage, 2009.
- [2] M. Diani and D. McAdam, *Social Movements and Networks: Relational Approaches to Collective Action*, ser. Comparative Politics Series. OUP Oxford, 2003.
- [3] M. Castells, *Networks of Outrage and Hope: Social Movements in the Internet Age*, ser. Polity Press. Wiley, 2012.
- [4] S. Gonzalez-Bailon, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, "The dynamics of protest recruitment through an online network," *Sci. Rep.*, vol. 1, 2011, 10.1038/srep00197. [Online]. Available: <http://dx.doi.org/10.1038/srep00197>
- [5] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, "Twitcident: Fighting fire with information from social web streams," in *Proceedings of the 21st International Conference Companion on World Wide Web*, ser. WWW '12 Companion. New York, NY, USA: ACM, 2012, pp. 305–308. [Online]. Available: <http://doi.acm.org/10.1145/2187980.2188035>
- [6] R. Li, K. H. Lei, R. Khadiwala, and K.-C. Chang, "Tedas: A twitter-based event detection and analysis system," in *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, April 2012, pp. 1273–1276.
- [7] M. Mathioudakis and N. Koudas, "Twittermonitor: Trend detection over the twitter stream," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 1155–1158. [Online]. Available: <http://doi.acm.org/10.1145/1807167.1807306>
- [8] T. Ahlqvist, M. Halonen, and S. Heinonen, "Weak signals in social media," *SOMED Foresight Report*, 2007.
- [9] C. Haythornthwaite, "Social network analysis: An approach and technique for the study of information exchange," *Library and Information Science Research*, vol. 18, no. 4, pp. 323 – 342, 1996.
- [10] S. L. Lim, D. Quercia, and A. Finkelstein, "Stakenet: Using social networks to analyse the stakeholders of large-scale software projects," in *Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering - Volume 1*, ser. ICSE '10. New York, NY, USA: ACM, 2010, pp. 295–304. [Online]. Available: <http://doi.acm.org/10.1145/1806799.1806844>
- [11] A. Rashid, A. Baron, P. Rayson, C. May-Chahal, P. Greenwood, and J. Walkerdine, "Who am i? analyzing digital personas in cybercrime investigations," *Computer*, vol. 46, no. 4, pp. 54–61, April 2013.
- [12] P. Taylor, C. Dando, T. Ormerod, L. Ball, M. Jenkins, A. Sandham, and T. Menacere, "Detecting insider threats to organizations through language change," *Law and Human Behavior*, vol. 37, no. 4, pp. 267–275, 2013.
- [13] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1 – 8, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S187775031100007X>
- [14] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 851–860. [Online]. Available: <http://doi.acm.org/10.1145/1772690.1772777>
- [15] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 492–499. [Online]. Available: <http://dx.doi.org/10.1109/WI-IAT.2010.63>
- [16] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu, "Predicting flu trends using twitter data," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, April 2011, pp. 702–707.
- [17] P. Rayson, "From key words to key semantic domains," *International Journal of Corpus Linguistics*, vol. 13, no. 4, pp. 519–549, 2008.
- [18] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in twitter events," *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, no. 2, pp. 406–418, Feb. 2011. [Online]. Available: <http://dx.doi.org/10.1002/asi.21462>
- [19] S. Prentice, P. J. Taylor, P. Rayson, A. Hoskins, and B. O'Loughlin, "Analyzing the semantic content and persuasive composition of extremist media: A case study of texts produced during the gaza conflict," *Information Systems Frontiers*, vol. 13, no. 1, pp. 61–73, 2011.