

21st ICCRTS

“C2 in a Complex Connected Battlespace”

Understanding Social Media Networks through Trending Analysis

Primary topic:

Topic 4 – Socio-technical Networks in Complex Connected Battlespace

Marie-Eve Jobidon & Bruce Forrester

Defence R&D Canada – Valcartier Research Centre

Point of Contact: Marie-Eve Jobidon

2459 de la Bravoure Road

Québec City, QC

G3J 1X5 Canada

Tel.: 418-844-4000 #4339

Marie-Eve.Jobidon@drdc-rddc.gc.ca

Abstract

The commander needs information superiority to achieve decision superiority, which ultimately leads to achieving desired effects from the employment of force. Information and decision superiority starts with shared situational awareness and understanding. This is achieved by first creating and maintaining a single authoritative source of data which reflects the needs of commanders and staff. The dramatic increase in information sources, and in particular the rise of social media as a source within open-source intelligence, has greatly complicated the ability to make sense from social sources. One valuable technique used to help manage the volume and variety of social media is trend analytics. Most social media platforms report trending data for the content being produced on their sites. However, exactly what counts as a trend is neither clear nor consistent between platforms. Are trends based on volume of posts, specific topics, words or hashtags, or sharing and similarity of content? What is the baseline timeframe that is used to determine a difference, and is there an actual statistical significance? Answers to these questions, as well as a good understanding of the nature of trends, are important for intelligence analysts in order to provide valid meaning from recognizing and analyzing social media trends for the commander.

Introduction

In military operations, the commander needs information superiority to achieve decision superiority, which ultimately leads to achieving desired effects from the employment of force. Information and decision superiority starts with shared situational awareness and understanding. This is achieved by first creating and maintaining a single authoritative source of data which reflects the needs of commanders and staff. The dramatic increase in information sources, and in particular the rise of social media as a source within open-source intelligence (OSINT), has greatly complicated the ability to make sense of information gathered from social sources. One valuable technique used to help manage the volume and variety of social media is trend analytics. Most social media platforms report trending data for the content being produced on their sites. For an intelligence analyst, challenges related to trend analysis are two-fold: 1) detecting or predicting trends and 2) determining what a given trend (or lack thereof) means from an intelligence and operational perspective; that is, what is the “so what” of the trend: Does it translate into behaviour or attitudes that can impact operational effectiveness? How can it be used to inform the commander’s decision making and contribute to mission success?

The rest of this paper is organized as follows. After a brief background on trends, we discuss how social media (SM) platforms define and use trends. In the second half of the paper, we describe a case study that explores how trend analysis can be exploited in the intelligence domain.

Trends

For our purposes, a trend can be defined as “a general development or change in a situation or in the way that people are behaving” (Cambridge English Dictionary, <http://dictionary.cambridge.org/dictionary/english/trend>). Trends represent the evolution of phenomena, and occur in virtually all aspects of life from the weather, the evolution of species, societies and culture, to economy, marketing,

fashion, food, art, and technology. They can be characterized by various natures, such as mathematical, psychological or behavioural. Among other factors, the nature of trends will impact how they manifest, and whether or how they are perceived by human beings. For instance, a mathematically significant trend can go unnoticed by people, either because it is not perceivable from a cognitive or psychological perspective or because it is not detected. Alternatively, we can perceive a trend that does not exist, such as when we detect a pattern in random data and mistakenly call it a trend. Indeed, the human brain tends to look for and detect patterns regardless of whether they exist or not, a psychological phenomenon referred to as apophenia [1, 2]. This cognitive predisposition has a mathematical equivalent in the form of the Type I error in statistics; that is, finding a significant result when none exist (the occurrence of which one wants to be as low as possible and easily identifiable).

Inherently, trends are time-based or time-bound. They can take place on various scales, be it centuries, generations, years and – as evident more recently with social media – even days, hours, and minutes. Time scale is critical in detecting trends through analysis. A trend is defined by its period of reference; that is, identifying a trend implies a comparison with a baseline, a time when that trend did not exist. In an intelligence context, the type of trends of interest and the way to approach, analyze, and exploit them may vary on whether the intelligence analyst's focus is, for instance, conducting a PMESII-PT¹ analysis to build a region profile or establishing situation understanding in reaction to a particular event such as a terrorist attack or a natural disaster (e.g., <http://www.sciencemag.org/news/2016/03/twitter-can-predict-hurricane-damage-well-emergency-agencies>). A time scale that is too short can miss longer term trends such as the slow rise in prices or a scale that is too long can miss the spikes in volume of SM posts (e.g., tweets) that occur around events. Hence, the nature of the phenomenon being sought, particularly the length, is critical to it being found.

Trend analysis consists in trying to determine whether an underlying pattern exists in data that was collected or that is streaming and constantly being captured. It can be used either to forecast future events or to better understand past or current events. The pattern (i.e., trend) is often hidden to various degrees in noise. This is especially the case with social and SM data due not only to the heterogeneous, unstructured, vast, and dynamic character of the data content but also to the diversity in the sources of data to be analyzed (various platforms, formats, etc.). For intelligence analysts, trend analysis is also challenging because they may not be aware a change is coming, making it difficult to know when to track something of relevance to the Commander. It is also tricky to determine what a SM trend actually means in the offline world in terms of situational understanding and prediction (i.e., the extent to which a trend will translate to changes of actions, behaviours or attitudes that can impact operations). By definition, a trend is identified as a significant change from a previous state. Knowing and understanding that previous state, in order to be able to identify the change, is important but can be challenging to achieve.

¹ Political, Military, Economy, Social, Infrastructure, Information systems, Physical environment, and Time.

Trends in Social Media

Several SM platforms offer a tracking functionality and information on what is considered “trending topics” (i.e., a general idea that is gaining popularity at a given time [3]). This section examines how trends are determined for several popular SM sites. Table 1 summarizes the various factors that are used by the trending models by the social media platforms described in more detail below.

Table 1. Comparison of characteristics that are used to determine trends in selected social media platforms.

Characteristic used in trend determination	Twitter	Google (+)	Facebook	Pinterest	YouTube
Topic	X	X	X	X	
Individual user engagement		X	X	X	
Population user engagement		X		X	X
Proprietary algorithm	X	X	X		
Timeliness	X	X	X		X
User determined location	X	X	X		X
Location of post					X
Overall trending topics in platform			X		
Things user liked			X		
List of followers	X				
Search volume		X			X
Language					
Age range					X

Note. There are also tools that track social shares and activity across these platforms in order to determine what content is trending.

Amazon trends purchase patterns and recommendations to encourage shoppers, and sites such as TripAdvisor trend search behaviour patterns in order to increase bookings and volunteer reviews. In addition, trends are combined into feedback mechanisms that encourage user interactions with many platforms as well as being publicly displayed for the same purpose.

Trending on Twitter

Twitter [4] trends are available for Apple and Android at twitter.com. “Trends are determined by an algorithm and, by default, are tailored for you based on who you follow and your location. This algorithm identifies topics that are popular now, rather than topics that have been popular for a while or on a daily basis, to help you discover the hottest emerging topics of discussion on Twitter that matter most to you” [4]. However, one can choose to see non-tailored trends by choosing a specific geographic location, which locates popular topics from the users at that location. Further, one will still see many

trending world conversations regardless of one's personal account settings. Twitter also offers email-based digests of popular videos and tweets trending in an individual's network.

Clicking on any of the displayed trends produces a page with the tweets that are making that topic trend. Previous trends are also available through a keyword search. Twitter also guards against abuse through active filtering or account suspension using the following rules [4]:

- Adding one or more topics/hashtags to an unrelated Tweet in an attempt to gain attention in search.
- Repeatedly Tweeting the same topic/hashtag without adding value to the conversation in an attempt to get the topic trending or trending higher.
- Tweeting about each trend in order to drive traffic to your profile or website, especially when mixed with advertising.
- Listing trends in combination with a request to be followed.
- Tweeting about a trend and posting a misleading link to something unrelated.

Apart from the above description, there is no indication of how the Twitter algorithm actually calculates the trends. In addition, advertisers and marketers can pay to add their subject or product to the list of trending topics. These paid trends get added to the top of the list and are specially marked to indicate sponsored content. Third-party applications exist that track trending topics, and "trending in Canada" for example is a mark of success in social media marketing for businesses operating or based in Canada.

Trending on Google and Google+

Google was one of the pioneers in trending starting with Flu Trends; now defunct. Each week, millions of users around the world search for online health information. As one might expect, there are more flu-related searches during flu season, more allergy-related searches during allergy season, and more sunburn-related searches during the summer. One can explore all of these phenomena using Google Trends. Google Flu Trends (Figure 1) launched in 2008 used the search results from populations in geological locations within the United States to predict the outbreak of the flu up to two weeks faster than the traditional systems.



Figure 1. Snapshot from Google Flu Trends. This site is now non-operational but shows that trending using automated means from socially produced data is not new.

Google Trends has moved beyond the flu. It now has six categories (business, entertainment, health, science/technology, sports, and top stories) and one can filter for 28 different countries. It compares a particular search-term to the total search-volume across various regions of the world and in various languages (Figure 2). “The horizontal axis of the main graph represents time, and the vertical is how often a term is searched for relative to the total number of searches, globally. Below the main graph, popularity is broken down by countries, regions, cities and language” [5]. Further, “Google Hot Trends is an addition to Google Trends which displays the top 20 hot – i.e., fastest rising – searches (search-terms) of the past hour in various countries. This is for searches that have recently experienced a sudden surge in popularity. For each of the search-terms, it provides a 24-hour search-volume graph as well as blog, news and web search results. Hot Trends has a history feature for those wishing to browse past hot searches” [5].

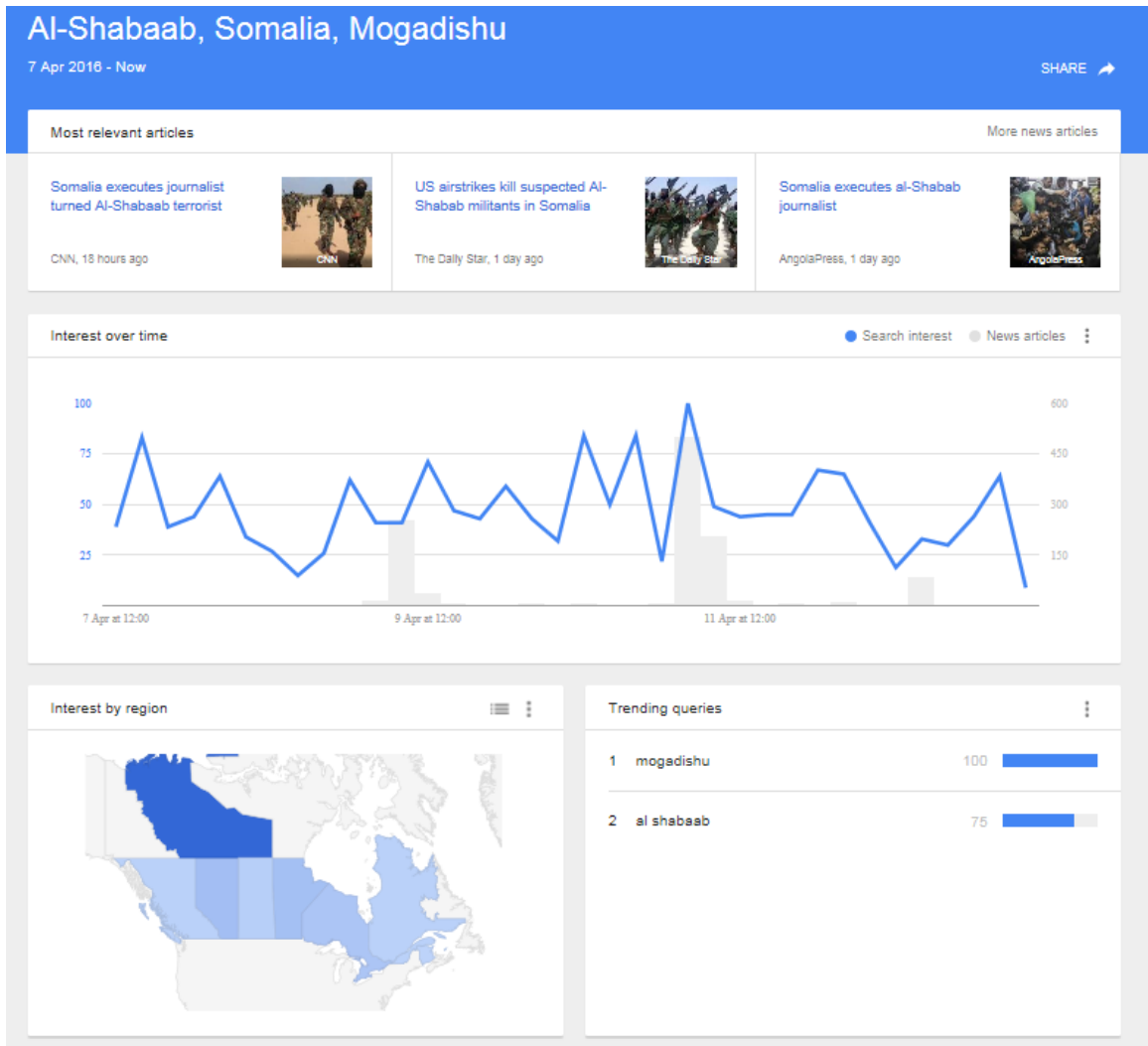


Figure 2. Snapshot from Google Trends on topic of Al-Shabaab [6].

Google + also has an entire web site dedicated to trending called “What’s Trending” that is dedicated to social media content (<https://plus.google.com/+WhatsTrending/posts>) .

Trending on Facebook

The trending pane shows up on the top right-hand side of the Facebook page and is currently only available in English in select countries. The trending pane displays a list of topics and hashtags that have recently spiked in popularity on Facebook [7]. However, contrary to what one might think, there is no common trending list across all Facebook members. The list and topics that are displayed are personalized based on a number of factors such as engagement, timeliness, pages you have liked, your location and what is trending across Facebook. On a computer, the topics are grouped into five categories: All News, Politics, Science and Technology, Sports, and Entertainment.

Apart from the above, there is no indication of how Facebook actually determines the trends; that is, there is no description of methods, metrics or time periods underlying the algorithm used to calculate trends.

Trending on Pinterest

Pinterest has a link dedicated to trends (<https://www.pinterest.com/explore/2016-trends/>). However, there is no information on how the photos become trends other than by “examining the pinning habits of more than 100 million users” [8]. For instance, in a post on the company blog about upcoming trends for 2016, they mention that they “looked at trending Pins in popular categories, Pins from our most influential Pinners and even a few hand-picked by Pinterest employees” in order to identify their predicted trends [9].

Trending on YouTube

YouTube has a trends dashboard (<https://www.youtube.com/trendsdashboard>) that allows one to select a US city or world country as well as specify an age group in order to filter the trends. The site allows the display of the “Most Shared” and “Most Viewed” videos, so apparently the trends are determined by simple counts of viewers. The site also has an interesting feature that allows the user to compare up to three sets of the filters described above. Users can also identify, when comparing, if trending videos are unique or common across the comparison filters (see Figure 3).

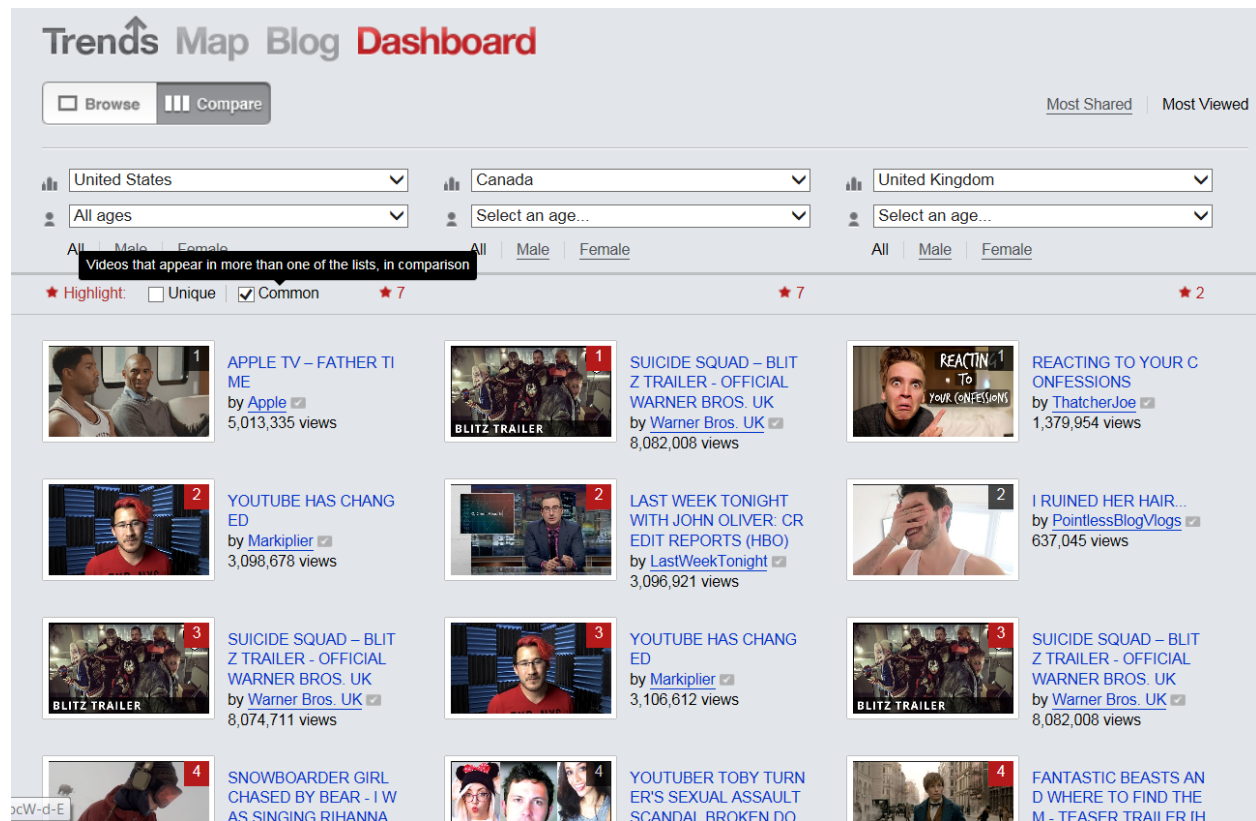


Figure 3. The YouTube trends dashboard that shows the comparison feature that also provides an indication of the number of videos in common.

Case Study

Framework

As is evident from the previous section, exactly what counts as a trend in SM is neither clear nor consistent across platforms. Little information is provided about how trends are determined. To what extent are trends based on volume of posts, specific topics, words, hashtags, or user interactions? What is the baseline timeframe that is used to determine that a trend is emerging, and is there an actual statistical significance? The validity or veracity of trends is also in question. For instance, what is the effect of media in the formation of trends? To what extent can trending topics be influenced by paying organizations or companies, or trends be “purchased”? Answers to these questions, as well as a good understanding of the nature of trends, is important for intelligence analysts in order to provide valid meaning from social trends for the commander.

In this paper, we present a case study in which we explore how SM trend analysis could be used in the intelligence context. Using NexaMaster (Nexalogy, Montreal, Canada), a social data analysis platform, we investigate ways in which an intelligence analyst could identify, track, and assess SM trends. The case study constitutes a post-hoc analysis of a situation that has trended in the past. Using some of the analytics tools currently available in NexaMaster, we examine the extent to which it is possible to identify the emergence of a trend, and track its evolution and nature (e.g., timeline, are official organizations or news outlets involved in the trend). We also assess how these SM trends correlated with the events that took place offline.

Case Study: An Analysis of #BlackLivesMatter

This case study is an examination of the events that occurred in Baltimore, Maryland, U.S., in April 2015 and the subsequent reactions on SM and offline. On 12 April 2015, Freddie Gray Jr. was arrested by the Baltimore Police Department (BPD) for possession of a switchblade knife; a weapon that the police alleged was illegal. Following his arrest, the 25-year-old African-American man was put in the back of a police van handcuffed (and later shackled) but unsecured, leaving him unable to protect himself from the movements of the vehicle and from slamming inside the van during transport. The van made four stops on the way to the police station (it is at the first stop that Freddie Gray, handcuffed when he was arrested, was also shackled). When the van arrived at the police station, Gray was unconscious and paramedics treated him for approximately 20 minutes before taking him to a trauma centre in a coma. He died from his injuries a week later, on 19 April 2015. The medical investigation showed the he sustained severe spinal cord injuries during transport that led to his coma and death. Table 2 presents the main timeline of events surrounding the death of Freddie Gray Jr. and its aftermath, which included several protests in Baltimore.

This event fuelled the Black Lives Matter movement on SM. The hashtag #blacklivesmatter was created in 2013 after a similarly racially-charged event in Florida, when a white man was acquitted of shooting of a black teenager. The movement was also fuelled, among others, by the killing of Michael Brown, an unarmed 18 year-old black man who was shot by a white police officer in Ferguson, Missouri, in August 2014.

Table 2. Timeline of the events surrounding the arrest and death of Freddie Gray Jr.

Date	Event
12 Apr 2015	Arrest and transportation unsecured in police van
18 Apr 2015	Protest outside of BPD
19 Apr 2015	Death of Freddie Gray Jr.
21 Apr 2015	Protests / 6 police officers suspended with pay pending investigation
23 Apr 2015	Protests in downtown Baltimore
24 Apr 2015	Statement by BPD commissioner acknowledging the arresting officers did not follow procedures
25 Apr 2015	Major protests in downtown Baltimore that turned violent
27 Apr 2015	Funeral and burial – Civil disorder intensifies in Baltimore (e.g., riots) State of emergency is declared and the National Guard is called in
29 Apr 2015	Protests in other U.S. cities in response to Freddie Gray Jr.’s death
30 Apr 2015	Medical Examiner reports that Gray sustained injuries as a result of slamming inside the van
1 May 2015	Charges filed against the 6 officers
3 May 2015	National Guard starts withdrawing and curfew is lifted
19 May 2015	Request for gag order on police, attorneys, and witnesses by the State’s Attorney (will be denied on 8 June 2015)
21 May 2015	Grand jury indictment of the 6 police officers

As can be seen in the table, protests and civil unrest lasted approximately two weeks. During that period, there was also substantial SM activity surrounding the events. In order to explore trend analysis in SM and its mapping to offline events, we used NexaMaster to collect and analyze SM data available around the time of the events, focusing on Twitter. First, a query was used in order to collect a sample of relevant tweets in the Baltimore area, using the following hashtags and words: #freddygray, #freddiegrey, FreddieGray, Freddiegrey, "Freddie Gray", "Freddie Grey", and #blacklivesmatter. This query yielded a collection of 39,302 English-language tweets. Then, category alerts were written for “Freddie Gray” and “protest”. A category corresponds to a pool of words related to a given topic (the category) that NexaMaster will search for in the collection of tweets. The analyst defines the categories and their content (the pool of words) depending on the question of interest. Alerts were triggered when the volume of occurrence for the event or category passed the two standard deviation mark. Each alert could then be compared retroactively to events in the real – offline – world.

Figure 4 shows the evolution in the volume of tweets for the sample collected, between 18 April and 4 May 2015. The number of tweets started increasing substantially following Gray’s death on 19 April, and peaked at various points in the following two weeks, on days where significant events occurred, such as protests on 23 and 25 April, and Gray’s funeral and ensuing civil unrest on 27 April. The peak on 1 May 2012 corresponds to the announcement of the six police officers involved being indicted of various criminal charges related to Gray’s death.

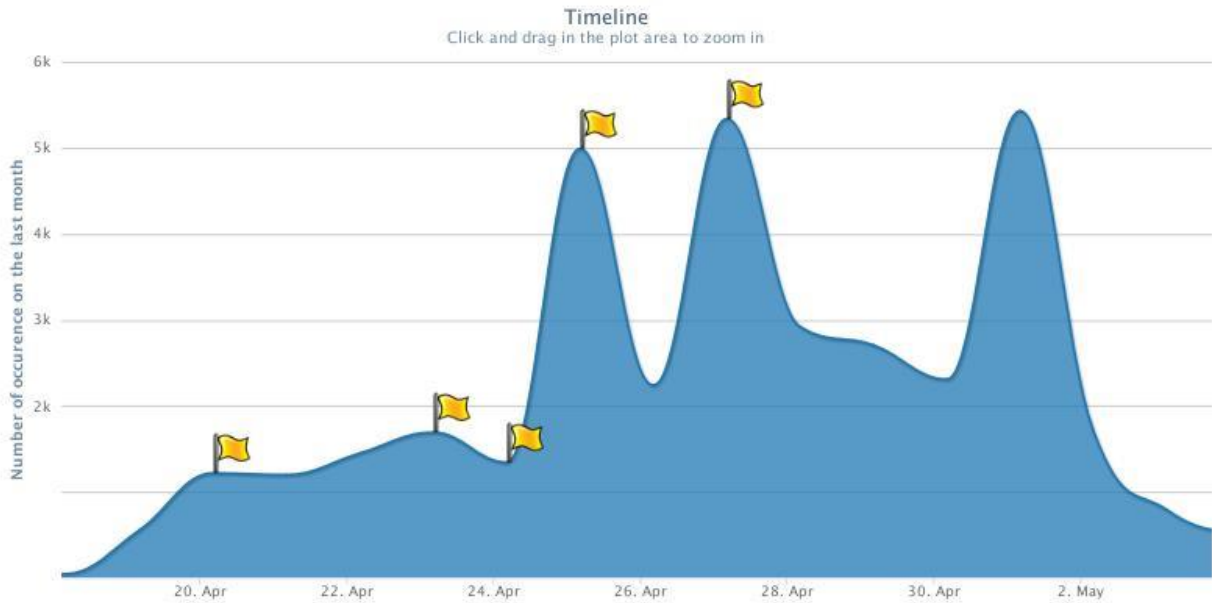


Figure 4. Timeline of volume of tweets with category peaks (indicated by the yellow flags).

Figure 5 illustrates the use of the words related to protesting (found in the “Protest” category) in the sample of tweets over time. The dashed lines identify days where significant events happened following the death of Freddie Gray Jr. Like Figure 4, this includes the major protests on 25 April and Gray’s funeral on 27 April. As can be seen in Figure 5, it is possible to identify trends related to protesting. That is, in the day or hours before protests took place, there was an increase in tweets containing words related to protesting. The red ellipses highlight two of these instances, prior to the protests on 23 and 27 April, respectively.

This suggests that it could be possible for an analyst tracking an area or topic of interest for the commander to detect brewing situations that could impact mission objectives, positively or negatively. The analyst could dig further in the content, characteristics, and demographics of the posts and publishers involved in the trend. For instance, looking at the tweets contained in the trend leading up to the protests on 23 April versus the ones at the peak on 23 April suggests that the proportion of tweets by media is greater at the peak than as the trend is developing.

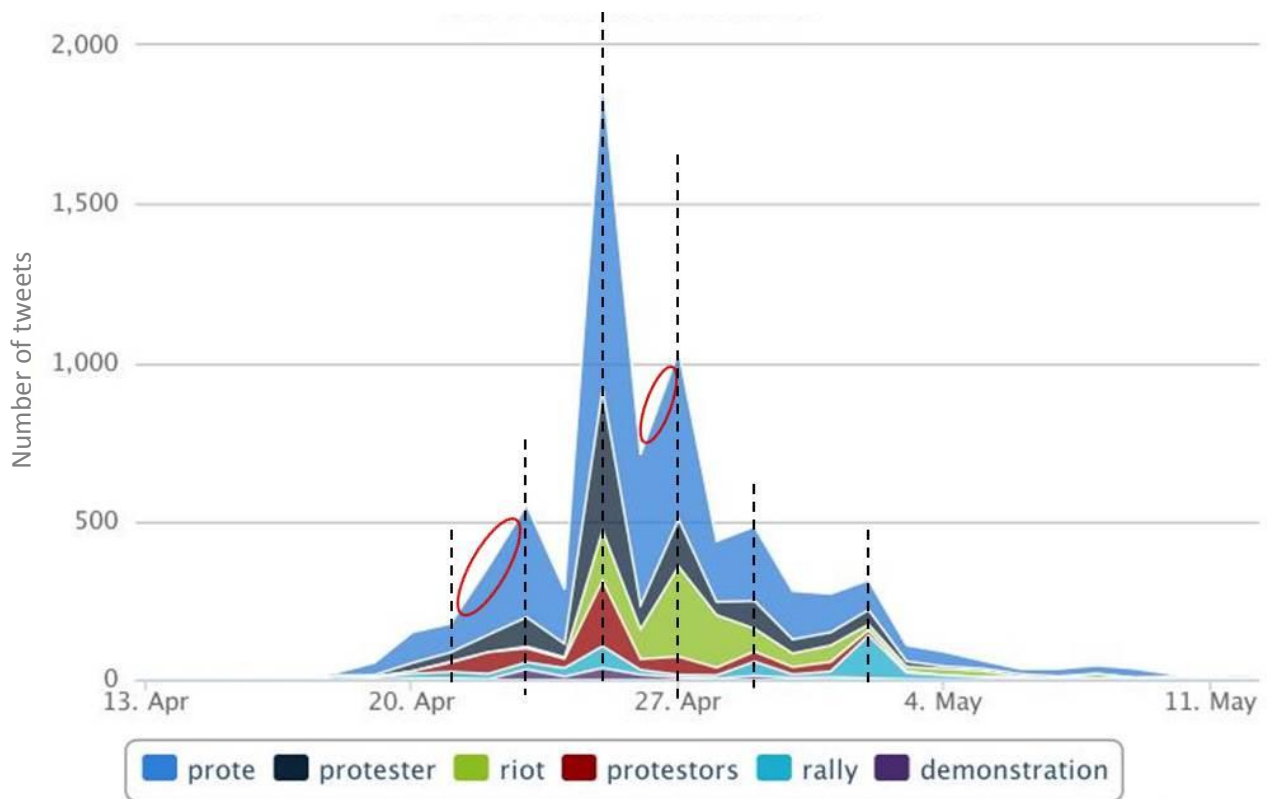


Figure 5. Timeline of the frequency of words related to protest. The dashed lines identify days where significant events happened following the death of Freddie Gray Jr. The red ellipses highlight two examples of trends that can be detected in that graph.

Discussion

The results of this early foray into trend analysis is promising, and should be explored and developed further. It must be remembered that we had prior knowledge of events that actually occurred in order to find the trends, so the next steps should include exploring the extent to which it is possible to detect or discover developing trends.

A number of factors can influence – or manipulate – SM trends and their evolution. That influence comes from the population contributing to the trend, but also from official organizations, media, celebrities, and bots that have an impact on trends through their contribution to the online “discussion”. Their influence can be more or less deliberate, and more or less well-intentioned, and can potentially amplify or smother a trend.

In addition, based on the information – or lack thereof – on how trends are detected and calculated on SM platforms, intelligence analysts should be wary of these trends since it is very difficult at this point to understand what they mean and to know how representative and valid they are. In order for trending topics to be useful, intelligence analysts need to understand how trends are identified by the

SM platforms that they are using as sources. An additional challenge with trends provided by SM platforms is that there is limited historical data available, which makes it difficult for an analyst to determine what topic have trended in an area of interest in the past and the evolution of these trends. All these aspects need to be taken into account when conducting trend analysis in an intelligence context. It is likely that trending tools will have to be designed and developed specifically for intelligence purposes in order to find trending of various lengths and sensitivities.

Conclusion

Trend analysis in SM is challenging for intelligence analysts not only because of the huge set of complex social data they have to work with, which makes identifying and tracking trends difficult, but also because it is challenging to determine the “so what”; that is, what a trend means with regards to situational understanding and predicting events. Ultimately, this is the knowledge that the intelligence analyst needs as one more source of information in order to best support the Commander.

Overall, the analysis of some of the current SM platforms and the case study lead us to the following conclusions:

- It seems possible to identify SM trends that correlate with offline events.
- The face validity of trending factors used by platforms seems good; however, the clarity of how trends are actually determined needs more definition and development.
- Intelligence analysts would not have a good level of confidence to place on trends from platforms with regards to population meaning, timescales used for baseline, magnitude and significance of change, etc.
- The representativeness of trends is in question.
- The fact that many platforms use individual preferences biases the trends; however, that might not be an issue for an analyst concentrating on specific regions.
- Analysts need to understand how each individual platform calculates trends. Alternatively, they may require trend analysis tools that are independent rather than relying on trending topics identified by platforms such as Twitter or Google+.
- A comparison feature, such as the one offered by YouTube, seems very useful for the exponential rise of online and embedded videos.

References

- [1] Brugger, P., *From haunted brain to haunted science: A cognitive neuroscience view of paranormal and pseudoscientific thought*. In J. Houran & R. Lange (Eds.) *Hauntings and Poltergeists: Multidisciplinary Perspectives*. In J. Houran & R. Lange (Eds.) *Hauntings and Poltergeists: Multidisciplinary Perspectives* 2001, NC: McFarland & Company Inc.
- [2] Fyfe, S., C. Williams, O.J. Mason, and G.J. Pickup, *Apophenia, theory of mind and schizotypy: Perceiving meaning and intentionality in randomness*. *Cortex*, 2008. **44**(10): p. 1316-1325.
- [3] McIntosh, H., *Trending topic*. In K. Harvey (Ed.), *Encyclopedia of social media and politics (pp. 1264-1265)* 2014, Thousand Oaks, CA: SAGE Publications.

- [4] Twitter, H.C. *FAQs about trends on Twitter*. 2016 [cited 2016 13 April]; Available from: <https://support.twitter.com/articles/101125#>.
- [5] Wikipedia. *Google Trends*. 2016 [cited 2016 13 April]; Available from: https://en.wikipedia.org/wiki/Google_Trends.
- [6] Google, t. *Google Trends - Al-Shabaab, Somalia, Mogadishu 7 April 2016 - Now*. 2016 [cited 2016 13 April]; Available from: https://www.google.ca/trends/story/CA_Ink_vh9XlgAwAADpOM_en.
- [7] FaceBook, H.C. *Trending*. 2016 [cited 2016 13 April]; Available from: <https://www.facebook.com/help/1401671260054622>.
- [8] Winterfeldt, M. *Pinterest announces the hottest home trends*. 2016# [cited 2016 14 April]; Available from: <http://www.popsugar.com/home/Pinterest-Predicts-Top-Home-Trends-2016-39401597#photo-39401597>.
- [9] Huynh, M. *Oh, How Interesting! The Pinterest 100: Fresh ideas for 2016*. 2015 [cited 2016 14 April]; Available from: <https://blog.pinterest.com/en/pinterest-100-fresh-ideas-2016>.