

#Ebola and Twitter: lessons learned



Tatiana Vorovchenko
Green Templeton College
University of Oxford

Dissertation submitted in fulfilment of the requirements for the
MSc by Research in Clinical Medicine

Word count: 29,933

Trinity term 2015

Contributions

This thesis would not have been possible without financial support of the ExxonMobil Foundation and Splunk4Good.

I used a Twitter dataset, which was provided by the generous purchase of the data by Splunk4Good. My own contribution includes the development of the RQs and design, all aspects of the literature review, analysis of Twitter data historic previews, development of filtering criteria for Twitter data retrieval and all analysis.

This thesis is based on my work at the Global Health Network, Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford. The research was completed in collaboration with Splunk4Good, the corporate social responsibility programme of Splunk Inc. Splunk team members assisted with the setup, installation and configuration of Splunk Enterprise 6.3.0, with data acquisition, retrieval and ingestion, and data model development.

The final dissertation was written by myself, with the invaluable comments and inputs from my supervisors, Dr. Francois Van Loggerenberg and Dr. Proochista Ariana. I bear responsibility for all the written work and reference of other authors.

#Ebola and Twitter: lessons learned.

Tatiana Vorovchenko, Green Templeton College

MSc by Research in Clinical Medicine

Trinity term, 2015

Abstract

Background: Twitter plays a role in information distribution during emergencies, and it is widely used by public health organisations during public health crises. Twitter has facilitated transmission of both accurate information and misinformation during the West African Ebola outbreak.

Study objective: To investigate the ways in which Twitter was used during the West African Ebola outbreak by the general Twitter population as well as by the World Health Organization (WHO) and U.S. Centers for Disease Control and Prevention (CDC) in order to better inform future efforts to harness the public health potential of Twitter.

Methods: The full historic Twitter dataset of over 60 million Ebola tweets from December 2013 to May 2015 was analysed using a platform for machine data analysis (Splunk Enterprise 6.3.0). In addition to descriptive statistics of Twitter use and engagement, content analysis was conducted to categorise Twitter accounts and tweets of interest.

Main results: The temporal tweet frequency distribution corresponded with key Ebola news events with a substantial proportion of tweets coming from West Africa. News, humour, Twitter celebrities', and celebrities' accounts achieved the highest level of engagement with the general Twitter population. The WHO and CDC also ranked high in engagement levels. Both the WHO and CDC actively used Twitter to communicate with the Twitter public. However, their strategies at disseminating public health information and guidance varied.

Conclusion: Twitter activity during the Ebola outbreak was more likely to coincide with news events rather than disease dynamics. Twitter provides a unique opportunity for public health organisations to listen to their audience and to share scientifically accurate information. Public health organisations could do more to engage with the Twitter population during the West African Ebola outbreak. Such organisations may benefit from employing humorous and emotional content and/or collaborating with accounts with a high level of Twitter engagement.

Table of contents

Contributions	2
Abstract	3
Table of contents	4
List of tables	6
List of figures	8
Acronyms	12
Chapter 1: Introduction	13
Chapter 2. Background. The role of Twitter in health: the case of Twitter and Ebola	15
2.1. Social media and its potential	15
2.2. Twitter	18
2.2.1. Twitter, its anatomy, and glossary of terms.....	18
2.2.2. Twitter usage	22
2.2.3. Twitter engagement.....	27
2.2.4. Twitter as a source of information for research	27
2.3. Twitter and public health	28
2.3.1. Twitter use by public health authorities	29
2.3.2. Twitter use by the WHO and CDC.....	30
2.3.3. Twitter use for disease detection and surveillance	32
2.3.4. Twitter use in emergencies.....	33
2.4. West African Ebola virus disease outbreak	34
2.5. Ebola on Twitter	38
2.6. Study rationale	43
Chapter 3. Methodological challenges of accessing and analysing Twitter data	45
3.1. Data sources, attributes, and costs	45
3.1.1. Twitter search option, Twitter search API, and Twitter Streaming API	45
3.1.2. Commercial vendors	46
3.1.3. Twitter data resellers	47
3.2. Sampling considerations and filtering	48
3.3. Data management and cleaning	49
3.4. Data analysis	50
Chapter 4. Methods	58
4.1. Data source	58
4.1.1. DataSift	58
4.1.2. DataSift historic previews	58
4.2. Sampling considerations and filtering criteria	59
4.3. Data ingestion and deployment	60
4.4. Data analysis on Splunk	61
4.5. Data cleaning	63
4.6. Description of the full dataset	64
4.7. Identification of study sub-periods	64
4.8. Description of study sub-periods	67
4.9. Retweets	68
4.10. Mentions and replies	72
4.11. Retweets, mentions, and replies of (inter)governmental organisation/NGO accounts	73

4.12. WHO and CDC Twitter Ebola activity	73
4.12.1. WHO and CDC Ebola tweets and their content	73
4.12.2. WHO and CDC retweets	75
4.12.3. WHO and CDC mentions and replies	76
4.12.4. Followers of the WHO and CDC accounts	76
4.13. Research ethics	77
Chapter 5. Results	78
5.1. Description of the full dataset	78
5.2. Study sub-periods and their correspondence with the Ebola infection rate and Ebola news events	79
5.3. Description of study sub-periods	82
5.3.1. Geography of tweets	85
5.3.2. Peak days in study sub-periods and URLs on these days	87
5.4. Retweets	97
5.5. Mentions	110
5.6. Replies	118
5.7. Retweets, mentions, and replies of (inter)governmental organisation/NGO accounts	122
5.8. WHO and CDC Twitter Ebola activity	124
5.8.1. WHO and CDC Ebola tweets and their content	126
5.8.2. WHO and CDC retweets	134
5.8.3. WHO and CDC mentions and replies	139
5.8.4. Followers of the WHO and CDC accounts	141
Chapter 6. Discussion and Conclusion	144
6.1. Pattern of Twitter use in the general Twitter population	144
6.2. Twitter engagement	148
6.3. WHO and CDC Twitter Ebola activity	156
6.4. Methodological contribution	160
6.3. Strengths and limitations	162
6.4. Wider implications of the study and lessons learned	167
6.5. Conclusion	172
Acknowledgements	174
References	175
Appendices	190

List of tables

Table 1. Key events of the West African Ebola outbreak.

Table 2. Fields of interest included in the project data model and their descriptions.

Table 3. Definitions of account types for Twitter account categorisation with examples.

Table 4. Categories of the WHO and CDC messages, definitions, and examples.

Table 5. Top 10 tweet text languages prevalent in the dataset.

Table 6. Description of the study sub-periods by length.

Table 7. Full description of the study sub-periods and the full period by the number of tweets, retweets, unique users, and URLs.

Table 8. Comparative description of peak days in each of the study sub-periods by the number of tweets, retweets, unique users, and unique links.

Table 9. Top 20 URLs shared on the 13 February 2014 (1st sub-period).

Table 10. Top 20 URLs shared on the 25 March 2014 (2nd sub-period).

Table 11. Top 20 URLs shared on the 8 August 2014 (3rd sub-period).

Table 12. Top 20 URLs shared on the 2 October 2014 (4th sub-period).

Table 13. Top 20 URLs shared on the 3 November 2014 (5th sub-period).

Table 14. The top 10 most retweeted tweets in the dataset.

Table 15. Retweet features of the 5 study sub-periods compared to the full study period.

Table 16. The number of tweets containing “any mentions”, mentions, and replies and their prevalence of the total number of tweets in the dataset.

Table 17. The number of unique mentioning and mentioned, replying and replied user accounts.

Table 18. Organisations owning (inter)governmental organisation/NGO accounts among the top 20 retweeted, mentioned, with the highest number of replies, or the top 10 that posted the most retweeted tweets.

Table 19. Description of the Twitter activity of the WHO and CDC during the Ebola outbreak.

Table 20. Categorisation of the WHO tweets.

Table 21. Categorisation of the CDC tweets.

Table 22. Distribution of the WHO tweets by type on peak days.

Table 23. Distribution of the CDC tweets by type on peak days.

Table 24. Top 10 most retweeted tweets posted by the @WHO account.

Table 25. Top 10 most retweeted tweets posted by the @CDCgov account.

Table 26. The most retweeted (more than 1,000 retweets within the first 3 days) WHO “Public health messages for the general public” tweets.

Table 27. The most retweeted (more than 1,000 retweets within the first 3 days) CDC “Public health messages for the general public” tweets.

Table 28. Ebola news events that made the general Twitter public turn to the WHO and CDC Twitter accounts.

List of figures

Figure 1. Number of social network users worldwide from 2010 to 2014, with projections to 2018.

Figure 2. Leading social networks worldwide as of August 2015, ranked by number of active users (in millions).

Figure 3. Correlation between Google Flu Trends and CDC surveillance data for the USA (2004-2014) and National Institute of Infectious Diseases for Japan (2004-2009).

Figure 4. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2015 (in millions).

Figure 5. Map of Twitter status object (tweet metadata) as of April 2010.

Figure 6. Demographic data for Twitter users among the USA online adult population.

Figure 7. Ranking of languages by the number of Twitter users.

Figure 8. The percentage of users self-reporting six of the most popular Twitter languages over time.

Figure 9. Language share of the 20 most active Twitter countries.

Figure 10. Precise location coordinates in the Twitter decahose 23 October 2012 to 30 November 2012.

Figure 11. Confirmed, probable and suspected EVD cases worldwide.

Figure 12. Temporal trends on Twitter and Google about Ebola and influenza (flu) before, during, and after Ebola cases in the USA, September to November 2014.

Figure 13. The percentage of tweets with different types of entities over time.

Figure 14. The percentage of tweets of different types over time.

Figure 15. The overall temporal tweet frequency distribution (by day) and the study sub-periods identified in the dataset.

Figure 16. The algorithm for Twitter account categorisation in the dataset.

Figure 17. Geographical distribution of tweets in the full dataset.

Figure 18. Overall tweet frequency distribution (by day) and the study sub-periods identified in the dataset.

Figure 19. Total reported cases of deaths from EVD over time.

Figure 20. New confirmed Ebola cases in Guinea, Liberia, and Sierra Leone during the West African Ebola Outbreak.

Figure 21. Geographical distribution of tweets in the 1st sub-period.

Figure 22. Geographical distribution of tweets in the 2nd sub-period.

Figure 23. Geographical distribution of tweets in the 3rd sub-period.

Figure 24. Geographical distribution of tweets in the 4th sub-period.

Figure 25. Geographical distribution of tweets in the 5th sub-period.

Figure 26. Temporal tweet frequency distribution (by day) during the 1st sub-period and the peak day.

Figure 27. Temporal tweet frequency distribution (by day) during the 2nd sub-period and the peak day.

Figure 28. Temporal tweet frequency distribution (by day) during the 3rd sub-period and the peak day.

Figure 29. Temporal tweet frequency distribution (by day) during the 4th sub-period and the peak day.

Figure 30. Temporal tweet frequency distribution (by day) during the 5th sub-period and the peak day.

Figure 31. Tweet and retweet temporal frequency distribution (by day) in the dataset.

Figure 32. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the dataset.

Figure 33. Tweet and retweet temporal frequency distribution (by day) during the full period.

Figure 34. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 1st sub-period.

Figure 35. Tweet and retweet temporal frequency distribution (by day) in the 2nd sub-period.

Figure 36. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 2nd sub-period.

- Figure 37. Time chart of the tweet and retweet temporal frequency distribution (by day) in the 3rd sub-period.
- Figure 38. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 3rd sub-period.
- Figure 39. Tweet and retweet temporal frequency distribution (by day) in the 4th sub-period.
- Figure 40. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 4th sub-period.
- Figure 41. Tweet and retweet temporal frequency distribution (by day) in the 5th sub-period.
- Figure 42. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 5th sub-period.
- Figure 43. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts within each of the 5 study sub-periods and over the full period.
- Figure 44. The proportion of different types of accounts that posted the most retweeted tweets within each of the 5 sub-periods and over the full period.
- Figure 45. Temporal frequency distribution (by day) of Tweets, “any mentions”, mentions and replies over the full study period.
- Figure 46. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the dataset.
- Figure 47. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 1st sub-period.
- Figure 48. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 2nd sub-period.
- Figure 49. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 3rd sub-period.
- Figure 50. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 4th sub-period.
- Figure 51. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 5th sub-period.
- Figure 52. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts within each of the 5 sub-periods and over the full study period.

- Figure 53. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the dataset.
- Figure 54. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 1st sub-period.
- Figure 55. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 2nd sub-period.
- Figure 56. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 3rd sub-period.
- Figure 57. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 4th sub-period.
- Figure 58. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 5th sub-period.
- Figure 59. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies within each of the 5 sub-periods and over the full study period.
- Figure 60. Temporal WHO and CDC tweet frequency distribution.
- Figure 61. Temporal distribution of retweets of the WHO and CDC Ebola tweets throughout the full study period.
- Figure 62. Number of WHO “Public health messages for the general public” retweets posted within the first three days of original tweet creation.
- Figure 63. Number of CDC “Public health messages for the general public” retweets posted within the first three days of the original tweet creation.
- Figure 64. Temporal frequency distribution of @WHO and @CDCgov mentions throughout the full study period.
- Figure 65. Time chart of the temporal frequency distribution of replies to the @WHO and @CDCgov throughout the full study period.
- Figure 66. Geographical distribution of @WHO mentions.
- Figure 67. Geographical distribution of @CDCgov mentions.
- Figure 68. Growth in the number of followers of the @WHO account over time.
- Figure 69. Growth in the number of followers of the @CDCgov account over time.

Acronyms

API	Application programming interface
CDC	U.S. Centers for Disease Control and Prevention
CSDL	Curated stream definition language
EVD	Ebola virus disease
GPS	Global positioning system
JSON	JavaScript object notation
NGO	Non-governmental organisation
RT	Retweet
RQ	Research question
SPL	Search processing language
UK	United Kingdom
URL	Uniform resource locator
USA, US	United States of America
UTC	Coordinated universal time
WHO	World Health Organisation

Chapter 1: Introduction

Twitter is an online social networking and micro-blogging platform that enables users to send and read short 140-character messages called "tweets". As of September 2015 Twitter had 316 million active users per month. (1) Twitter is used by different actors in health sector to varying degrees, and attracts the interest of researchers due to data availability, multiple authorship, agent interaction, temporal dynamics, instantaneity, and ubiquity. (2)

There is a growing body of research examining the role of Twitter in public health. It has been demonstrated that Twitter plays a role in information distribution during emergencies, and that it is widely used by public health organisations during public health crises. (3) Existing evidence indicates that Twitter has played a role in information transmission during the Ebola outbreak in West Africa that began in December 2013 and that Twitter conversation about Ebola was initiated and amplified by the Western media coverage. (4–10) Governmental health organisations that are involved in the outbreak response at the global scale, such as the World Health Organisation (WHO) and US Centers for Disease Control and Prevention (CDC) use Twitter as one of their official communication tools. However, the literature on Twitter engagement related to this outbreak and the ways in which global health organisations, such as the WHO and CDC used the platform during this period have not been explored.

The objective of this thesis is to investigate the ways in which Twitter was used during the West African Ebola outbreak by the general Twitter population as well as by the WHO and CDC. I will be addressing the following research questions (RQs):

RQ 1. What was the pattern of Twitter use in the general Twitter population during the West African Ebola outbreak?

RQ 2. How engaged were Twitter users in the Twitter Ebola conversation? What accounts demonstrated the highest level of engagement on Twitter related to the West African Ebola outbreak? How did governmental and non-governmental health organisations fit in?

RQ 3. How did the WHO and CDC use Twitter during the West African Ebola outbreak? What level of engagement did the WHO and CDC achieve with the general Twitter population in relation to the West African Ebola outbreak?

In this thesis, I outline the ways Twitter is used for health and the role (positive and negative) Twitter has played in the West African Ebola outbreak. I identify the gaps in our understanding of how Twitter users have engaged on Twitter in relation to this outbreak (chapter 2). I then review methodological challenges of health research on Twitter (chapter 3), and elaborate the methods used in my research (chapter 4). I present my results (chapter 5) and discuss these findings in the context of past research and future needs before concluding with acknowledgement of the limitations of my study and suggestions for future investigation (chapter 6).

Chapter 2. Background. The role of Twitter in health: the case of Twitter and Ebola

2.1. Social media and its potential

As social media continues to evolve, its definition is constantly changing to capture all its features and reflect the role it plays in the modern world. According to Cohen, (11) social media "are the platforms that enable the interactive web by engaging users to participate in, comment on and create content as means of communicating with their social graph, other users and the public". Social media continues developing and integrating deeply into human lives and may serve a variety of purposes such as social interaction, information seeking, time passing, entertainment, relaxation, communicatory utility, expression of opinions, convenience utility, information sharing, and surveillance and watching others. (12) For example, LinkedIn allows its users to build professional connections, Facebook is widely used to connect with friends, Twitter allows public sharing of short messages, Instagram is used to share favourite pictures, and YouTube allows the sharing of videos.

The number of active social media users has been growing rapidly. As of 2015 it is estimated that nearly 2 billion people globally use social networks (Figure 1). Social media platforms have differing levels of popularity and a number of active users. As of August 2015 Facebook is the most popular platform with 1,490 million users (Figure 2).

Figure 1. Number of social network users worldwide from 2010 to 2014, with projections to 2018. (13)

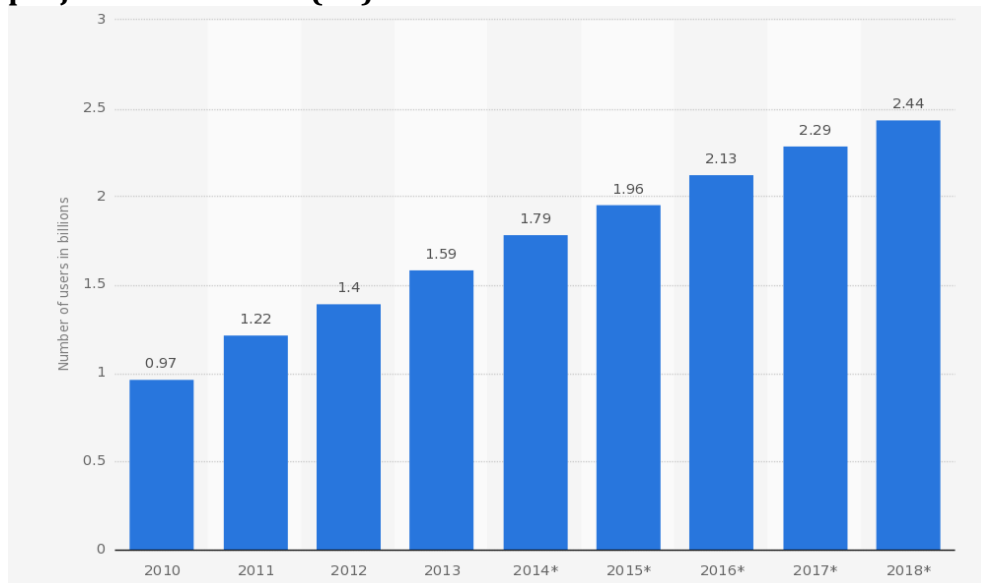
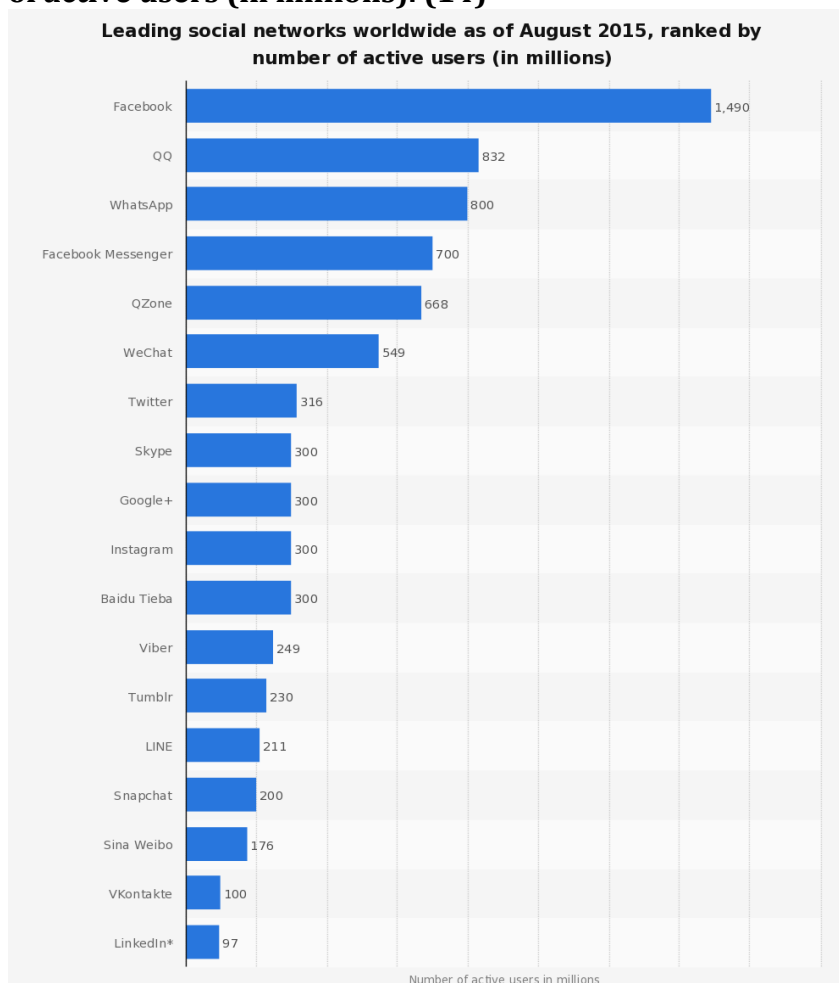


Figure 2. Leading social networks worldwide as of August 2015, ranked by number of active users (in millions). (14)



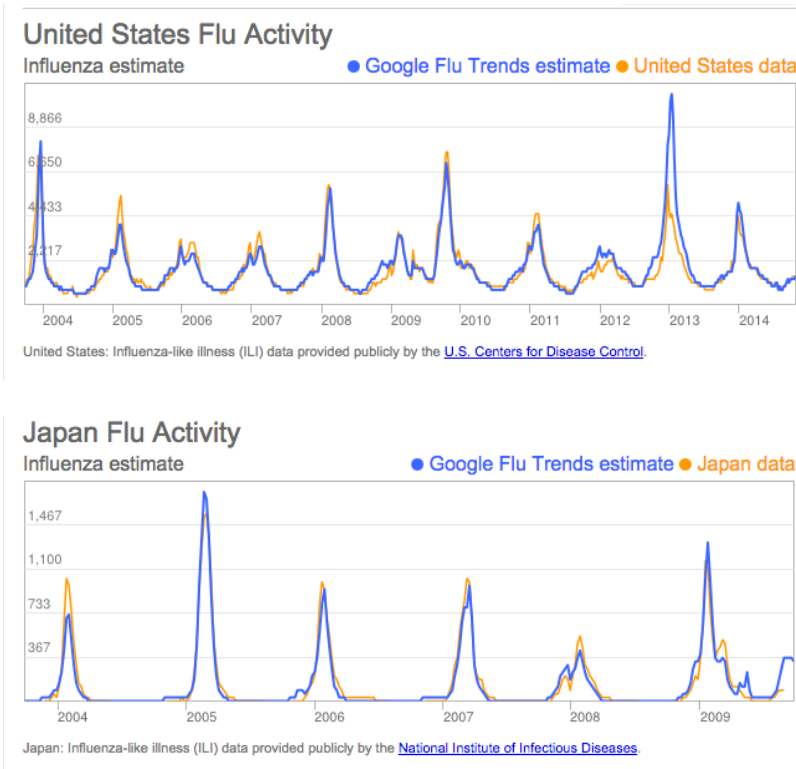
Social media platforms contain a huge amount of information. The rate of new data produced in Facebook is 600 terabytes¹ daily. (15) At the end of 2013, around 350 million new photos were uploaded on Facebook daily and around 3 billion likes occurred every day. (16) 500 million tweets are sent daily (1) and over 100 hours of video are uploaded on YouTube every minute. (17) Therefore, social media platforms are undoubtedly sources of large datasets, known as Big Data.

Big Data is a broad term for the datasets that are distinguished by their size or complexity. (18) Traditional data processing applications are inadequate to capture, store, and manage them. Big Data is already being used in various facets of our life. For example, Google uses the data of high-resolution maps of the world to develop a driverless car; Amazon uses the data of its user preferences and purchases to offer their customers products that they will be more interested in. (18) Big Data is also being used for health and human welfare. One example of this is Google Flu Trends². Millions of users around the world search for health information online. Google estimates how much flu is circulating in different countries around the world using the data of particular search queries on its search engine. (19) These data correlate with the data from traditional flu surveillance systems (Figure 3). (20) The reporting lag of these predictions is around one day, whereas traditional surveillance systems might take weeks to collect and report the data.

¹ 10¹²bytes or 1,000 GB

² <https://www.google.org/flutrends/>

Figure 3. Correlation between Google Flu Trends and CDC surveillance data for the USA (2004-2014) and National Institute of Infectious Diseases for Japan (2004-2009). (21)



In summary, big datasets can provide valuable insights. Social media is also a source of Big Data that may be both historic and real-time data. These data are global, in some cases public and readily available for analysis and, moreover, can be reanalysed for exploration of newly emerging research questions.

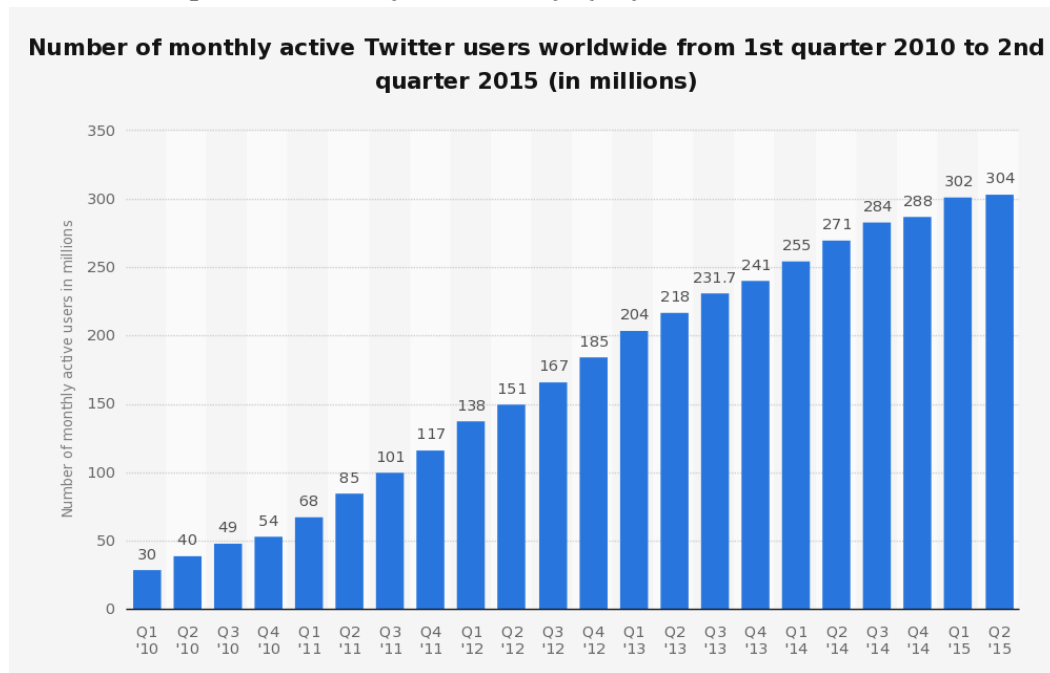
2.2. Twitter

2.2.1. Twitter, its anatomy, and glossary of terms

Twitter is an online social networking and micro-blogging platform that enables users to send and read short 140-character messages called “tweets”. Micro-blogging allows users to exchange small elements of content: short sentences, individual images, or video links. (22) Twitter is currently primarily an online service accessible from computers, tablets and mobile phones. Since its launch in 2006, the population of Twitter

users has been constantly growing (Figure 4) and as of September 2015 had 316 million active users per month. (1)

Figure 4. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2015 (in millions). (23)



Each Twitter user has a profile, which displays publicly available information, as well as all the tweets they have posted. In their profile each user might include a biography (up to 160 characters), which is a user description. It is also possible to provide user location information, a link to a personal website, and the date a user joined Twitter. Some of the user accounts can be “verified” by Twitter³. Twitter users may “follow” each other’s accounts, therefore, becoming a “follower” and having “followers”. A user may “favourite” a tweet to indicate that they liked a specific tweet and they may forward a tweet to their followers using Twitter’s retweet feature. Tweets may contain a variety of forms of information, such as text, hashtags, mentions, photos, videos, and URLs⁴. A hashtag symbol “#” helps to tag a keyword in a tweet. A mention symbol “@” indicates a reference to another user or a direct reply to them. An URL is a web address that points to a unique page on the Internet. A “search” box provides an opportunity to search all public tweets for

³ “Verification is a process whereby a Twitter account receives a blue check that establishes authenticity of identities of key individuals and brands on Twitter”. (24)

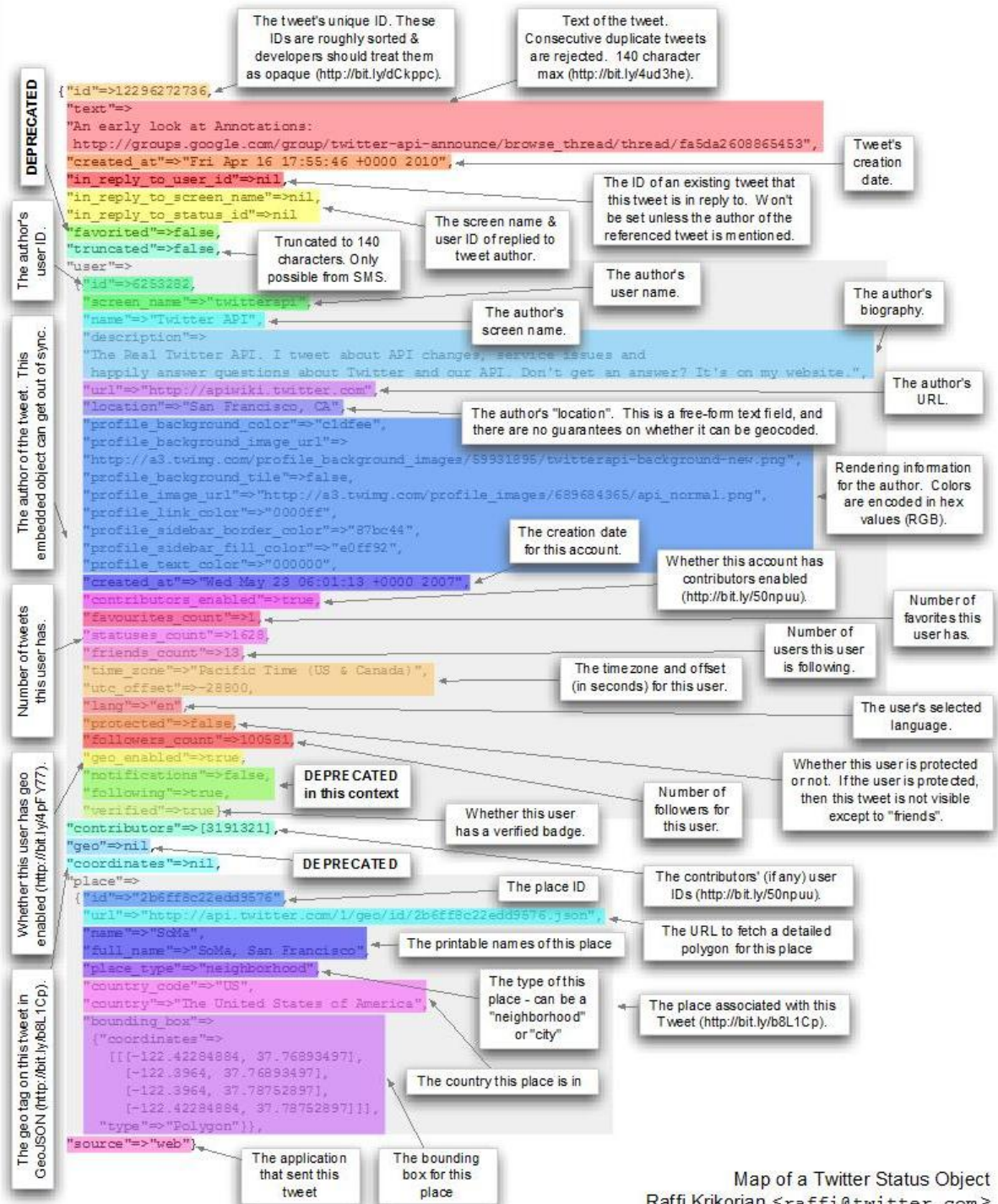
⁴ Uniform Resource Locator

keywords, usernames, hashtags, or subjects. The same can be performed using search.twitter.com. The full Twitter glossary of terms is presented in the Appendix 1. All the described information can be retrieved with a tweet as its metadata, which shows that a tweet contains much more information than just a 140-character text message (Figure 5). The Twitter streams use JSON⁵ as the standard format.

Not all the tweets that have been posted remain publicly available. There are four mechanisms that might make a public tweet unavailable: a tweet could be deleted by a user; users might make their accounts visible only to their followers; an account can be suspended by Twitter; and a user can deactivate their account. (25) As of January 2014, up to 20% of publically posted tweets since July 2009 had become unavailable. (25)

⁵ JSON stands for JavaScript Object Notation, and is a text-based standard for data interchange.

Figure 5. Map of Twitter status object (tweet metadata) as of April 2010. (26)



2.2.2. Twitter usage

2.2.2.1. Demographics of Twitter users

The data on demographic characteristics of Twitter users, such as gender, socio-economic status, and age is limited, and present in the literature only for some particular countries. The demographic profile of the USA Twitter users, where about 23% of registered Twitter accounts come from, (1) are well studied. According to the survey from the Pew Research Center⁶, in September 2014, among 81% of American adults who are online⁷, 23% used Twitter (24% of men and 21% of women). (27) Moreover, it has been shown that Twitter users tend to be younger, have higher levels of education and socioeconomic status, and live in urban areas (Figure 6). Therefore, it is clear that the Twitter population does not represent the whole population in one particular country or at the global scale and the findings of research using Twitter data should apply only to the population of Twitter users.

⁶ The survey data is based on the telephone interviews with a nationally representative sample of 2,003 adults living in the continental United States.(27)

⁷ 18-year-old and over, who say that they use Internet or e-mail.

Figure 6. Demographic data for Twitter users among the USA online adult population. (27)

Twitter users

Among online adults, the % who use Twitter

	2013	2014
All internet users	18%	23%*
Men	17	24*
Women	18	21
White, Non-Hispanic	16	21 *
Black, Non-Hispanic	29	27
Hispanic	16	25
18-29	31	37
30-49	19	25
50-64	9	12
65+	5	10*
High school grad or less	17	16
Some college	18	24
College+ (n= 685)	18	30*
Less than \$30,000/yr	17	20
\$30,000-\$49,999	18	21
\$50,000-\$74,999	15	27*
\$75,000+	19	27*
Urban	18	25*
Suburban	19	23
Rural	11	17

Source: Pew Research Center's Internet Project September Combined Omnibus Survey, September 11-14 & September 18-21, 2014. N=1,597 internet users ages 18+. The margin of error for all internet users is +/- 2.9 percentage points. 2013 data from Pew Internet August Tracking Survey, August 07 - September 16, 2013, n= 1,445 internet users ages 18+.

Note: Percentages marked with an asterisk (*) represent a significant change from 2013. Results are significant at the 95% confidence level using an independent z-test.

PEW RESEARCH CENTER

2.2.2.2. Twitter languages and their geography

The Twitter platform currently supports more than 35 languages, (1) whereas at least 78 languages were identified in tweets in 2010-2012. (28) Not surprisingly, English is the most dominant language among Twitter users, and Spanish comes second (Figure 7). (28) The number of users reporting English as their language (which may differ from the actual language they tweet in) has been decreasing over time. In January 2010 83% of users reported English, while in December 2013 52% of them did (Figure 8). (25) At the same time increase in a large number of other languages, such as Spanish, Japanese, and Portuguese was observed, suggesting that the Twitter population is continuing to become more global and diverse. (25) It is worth noting that Twitter languages do not correlate with the distribution of languages around the world, according to which Chinese ranks first, followed by Spanish, English, and Hindi, (29) reminding us that Twitter users are a

specific population of micro-blogging platform users. Moreover, in some countries, such as China, North Korea and Iran, Twitter is blocked. (30)

Figure 7. Ranking of languages by the number of Twitter users. (31)

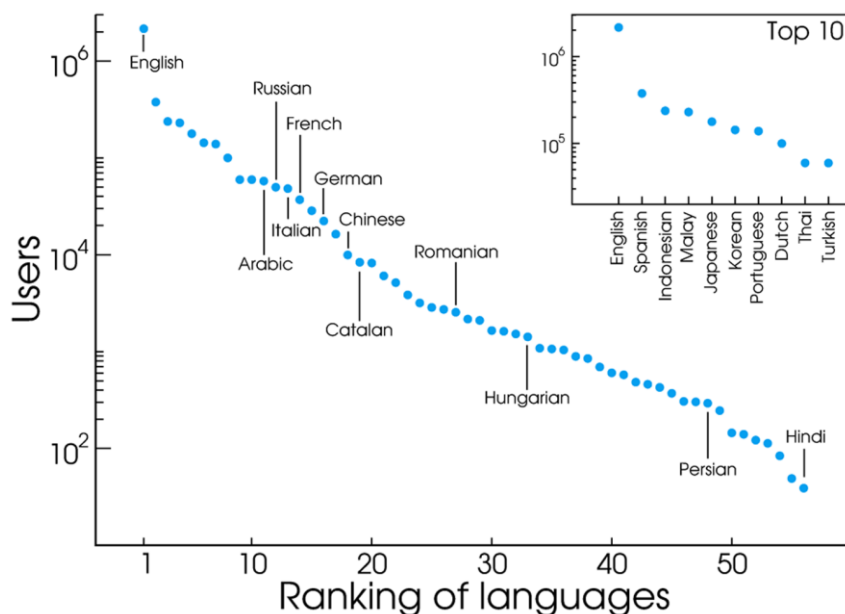
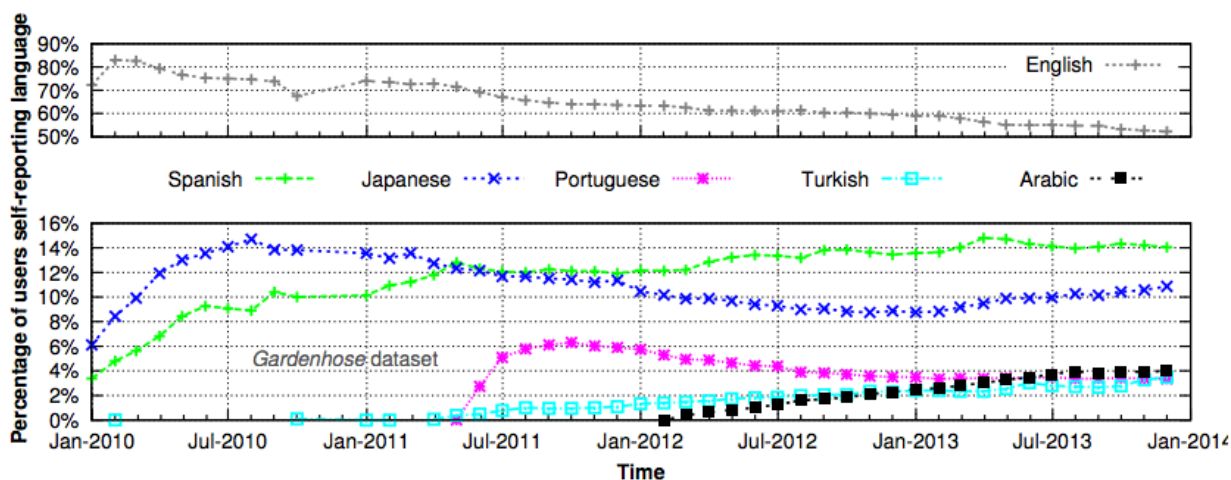


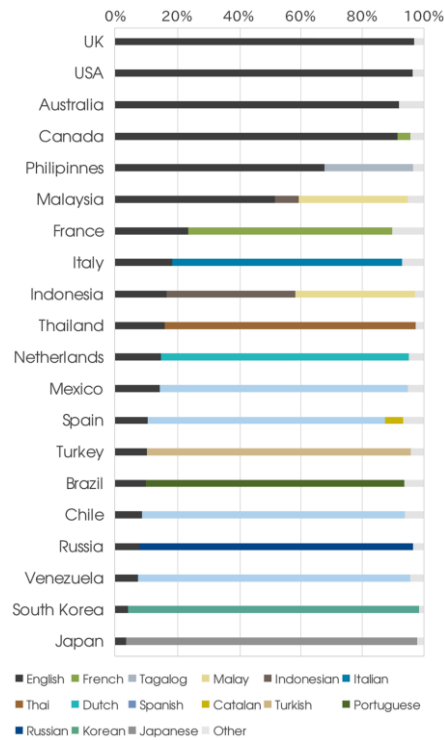
Figure 8. The percentage of users self-reporting six of the most popular Twitter languages over time. (32)



English is plotted separately in the top graph in order to increase the readability.

Figure 9 demonstrates the proportion of languages adopted by users for tweeting in top 20 Twitter active countries. It can be seen that English is widely used by Twitter users even if it is not their native language, partly in order to reach a wider audience. (28)

Figure 9. Language share of the 20 most active Twitter countries. (33)



2.2.2.3. Geography of Twitter users

There are two types of geolocation information available on Twitter: place, such as city or neighbourhood (manually selected by a user from a predefined list of locations supported by Twitter and primarily used for tweeting from fixed-location devices) and precise location (a set of coordinates provided by GPS⁸). (34) Once a user tweets with a place location, Twitter will automatically attach a general location label to the next tweets. For example, Twitter would attach the last selected location to users traveling to another country unless they manually change it. Using Twitter for Android or Twitter for iOS⁹ (on mobile devices) users may “geo-tag” a tweet indicating their precise geographical location at the time the tweet is sent, which will include longitude and latitude coordinates. It means that users do not have to update their location manually. (34) However, this option

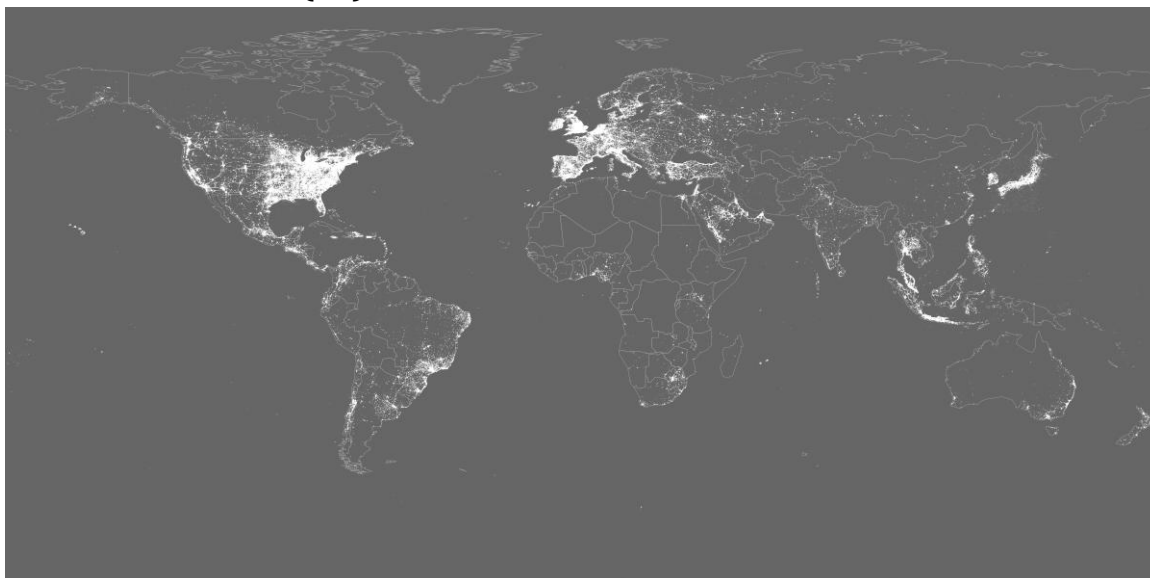
⁸ Global Positioning System

⁹ Android and iOS are mobile operating systems.

is disabled by default and users must explicitly alter the setting to enable it. In addition, Twitter allows third-party applications or websites to attach user's precise location. (35)

A comprehensive study of a large dataset (over 1.5 billion from Twitter decahose¹⁰) from 23 October 2012 to 30 November 2012 demonstrated that on a typical day, 1.8% of tweets had a place indicator and 1.6% of tweets had a precise location. (36) Figure 10 demonstrates precise location coordinates of these 1.6% geo-tagged tweets. Most of the tweets were posted from the USA (in North America), Western Europe (in Europe), Brazil and Argentina (in Latin America), Philippines, Malaysia, Indonesia, and Japan (in Asia), and Middle Eastern countries, while in Africa most entries came from Nigeria and South Africa. Another study demonstrated that in July 2014 among 1.2% of geo-tagged tweets, 32% of tweets came from the USA and Canada, tweets coming from Latin America, Asia, and Europe had a prevalence of 20% from each of these locations, and approximately 10% of tweets came from the Middle East. (25)

Figure 10. Precise location coordinates in the Twitter decahose 23 October 2012 to 30 November 2012. (37)



¹⁰ Random 10% of tweets.

2.2.3. Twitter engagement

The term “Twitter engagement” can be understood as the interactive, synchronous communication and collaboration among numerous participants via Twitter. (38) Twitter is widely used as a communication channel across a wide range of disciplines, from politics, news, sports, and marketing to crisis communication, as well as for interpersonal communication. Twitter is based around a broadcast communication model, where users publish tweets to be read by all other users of the service. (36) In turn, organisations and individuals subscribe to other Twitter users to receive their updates. Twitter enables communication with delays of milliseconds regardless of distance, and offers researchers a unique opportunity to study social interactions and engagement of the general Twitter population at a global scale. Twitter users may engage with other users on Twitter by retweeting their tweets, by mentioning them, sending them direct replies, or favouring their tweets. These parameters can be used for Twitter engagement measurement.

2.2.4. Twitter as a source of information for research

Twitter as well as other social networks have attracted the interests of researchers due to several particularities they have: data availability, multiple authorship, agent interaction, temporal dynamics, instantaneity, and ubiquity. (2) Firstly, it is a source of huge amount of data, which is constantly generated and might be freely available to researchers (with some limitations and privacy considerations). This volume of data makes it possible to study Twitter social data in the context of Big Data research. Secondly, the data are produced by a large number of diverse authors distributed globally. Thirdly, users interact with each other and take personal decisions, which makes it possible to study human behaviour and interactions. Fourthly, as data are generated continuously, it is

possible to study spatio-temporal processes and changes, looking at topic evolution and social mobilisation during real-world events. Moreover, as the data are continuous, social networks react in real time to any internal or external stimuli. Finally, Twitter data can be generated anywhere at any time due to prevalent Twitter use from mobile devices. Moreover, as some of the data are geo-tagged, it is possible to analyse geographical patterns. Additionally, Twitter has well-defined interface for software developers¹¹ that makes it possible to extract the real-time data. (2) Twitter is considered to be the most studied social network. (39)

2.3. Twitter and public health.

The general public, hospitals, governmental, non-governmental, and philanthropic health organisations, the private health sector, and civil society use Twitter to varying degrees in order to communicate health information. Not surprisingly, researchers from around the globe are interested in investigation of Twitter and health-related issues, as well as private and government-funding agencies that provide grants for research in this area. (40)

In September 2012, 72% of Internet users in the USA, the country which has the biggest proportion of Twitter accounts, said that they looked online for health information within the past year, (41) 42% of consumers in the USA reported that they use social media in order to access consumer reviews of treatments, hospital and physicians, 25% posted about their health status at least once, and 20% joined a health-related community or forum. (42) Moreover, 80% of individuals in the age group 18-24 years old were likely to share their health information online, while 90% would engage in health activities or trust health information found on social media. (42)

¹¹ <https://dev.twitter.com>

I will describe further how Twitter is being used in public health by public health authorities with the emphasis on the WHO and CDC, as well as how the platform might be used for disease detection and surveillance and in public health emergencies.

2.3.1. Twitter use by public health authorities

Twitter use by organisations is increasing significantly. (43) They use it for brand enhancement, increased visibility, customer support, networking, internal communication, lead generation, and other online presence support. (80) Non-profit organisations are also present on Twitter, which they use in order to promote the organisation and mobilise their audience. (45) Social media use by health organisations is growing in both private and public sectors. (46,47) Ministries of health around the world widely use Twitter. (48) It is one of the most commonly used social media channels adopted by the US state and local health departments. (49–51)

Twitter is used for a variety of reasons in the public health field. Firstly, it is used for the dissemination of health information and health promotion. For example, the US local health departments communicated diabetes information to the public on Twitter, providing information on the diabetes risk factors, specific cues to action, and benefits of a healthy behaviour. (52) Public health authorities and hospitals have also used Twitter for promotion of smoking-cessation services (53–55) and for sexual health promotion. (56)

Secondly, public health departments may use Twitter in order to identify public health problems in the area. For example, the Chicago health department in the USA used Twitter in order to identify and respond to complaints about possible food-borne illness, providing a link to fill out a formal complaint form. This intervention was followed by an inspection of the mentioned restaurants and helped to identify the places that did not meet the safety criteria. (57)

Thirdly, Twitter is an important tool in the health policy landscape. By supporting the conversation between people from different communities and professions, it gives insights to public health authorities about people's opinion and sentiment about health policies that are either being prepared or have been put into place. For example, Twitter response to recommendations against screening with prostate-specific antigen in the USA was explored. (58) This study found a significant increase of tweets on the prostate-cancer topic after the release of a corresponding policy. The negative feeling towards English National Health Service reforms in 2011 (Health and Social Care Bill) on Twitter also reflected the results of the national polling data. (59)

2.3.2. Twitter use by the WHO and CDC

The WHO and CDC are among the health organisations operating globally that have social media teams and accounts on Twitter and actively use them to disseminate health information, counter rumours and engage with their audience.

WHO started to use social media based on the lessons learned during the swine flu pandemic response in 2009, when a lot of rumours circulating on social media were missed by the WHO. Having realised that, the organisation changed its approach to social media and in October 2011 it had 255,000 followers on its Twitter account (@WHO). (60) As of September 2015 this number had reached 2.64 million. One of the successful examples of Twitter use by the WHO was in March 2011. Three days after the tsunami hit Japan, WHO noticed that some people were drinking iodised wound cleaner and consuming large quantities of iodised salt, which was thought to be protective from radiation leak at the Fukushima nuclear plant (rather than recommended potassium iodine tablets). (61) Both misconceptions were immediately addressed via social media channels including Twitter, the medium that helped to reveal the problem and tackle it. A few days later it was

reported that Chinese people were trying to return their salt, indicating that the message was successfully conveyed. (61)

The CDC encourages its staff and departments to strategically use Twitter in order to effectively and inexpensively reach individuals and stakeholders with timely health and safety information. It has more than 50 accounts on Twitter, representing campaigns, programmes, diseases and conditions, and emergency preparedness among others; moreover, the CDC directors also have their professional Twitter accounts. (62) As of September 2015 the @CDCgov Twitter account had approximately 586,000 followers. Since the adoption of Twitter by CDC it has been noticed that the public is increasingly relying on social media for health information. For example, during the 2009 influenza (H1N1) outbreak the number of followers of CDC Twitter account (@CDCemergency) increased from 65,000 to 1.2 million within a year. (3)

CDC identified the following characteristics of social media that make it a viable tool for public health information communication: it can work best when integrated with traditional public health communication channels; it is possible to target and reach diverse audiences; it allows sharing of public health information in new spaces; it is possible to listen and collect feedback in real-time; and it enables direct engagement. (63)

The following principles were proposed for social media engagement for public health communications: listening to social media conversations; engaging with influencers and their conversations; responding to questions and comments received through social media channels; creation of opportunities for users to engage with the organisation and with each other; welcoming user-generated content; creation of opportunities to integrate online and offline engagement; and leveraging social media for community engagement. (38)

2.3.3. Twitter use for disease detection and surveillance

Twitter conversations are of interest for public health authorities and researchers due to their potential to provide data for prediction and tracking epidemics, as it is being done with the Google Flu Trends. The general idea is that the frequency of tweets on a particular topic approximates interest in a given public health topic. There is a big body of research that attempts to evaluate the ability of Twitter data to track and predict influenza, (64,65) norovirus, (66) dengue (67) and other disease outbreaks. (68) While some studies used retrospective data, (69,70) other demonstrated the ability of Twitter to predict epidemics in the real time. (71,72) It has been shown that Twitter data may substantially improve the prediction of emergence and spread of influenza in the USA in addition to the CDC models (69,70,73–76) and in the UK. (77,78)

However, the ability of Twitter to predict and track infectious diseases depends on several factors, among which are the Twitter penetration in a studied location, public awareness about the disease of interest and its risk factors, social perception of the disease, the age group the disease affects, and the ways the disease is measured. It is also important to account for seasonal factors that may influence the frequency of tweets on a particular topic. Twitter conversation might also be highly influenced by contextual factors and high media interest, as it has been demonstrated on the case of Ebola communication. (79) It can also be challenging to choose the key words and hashtags that should be included in filtering criteria for some public health topics. For example, it can be easier to define filtering criteria for influenza and insomnia, but more challenging for cardiovascular diseases and conditions such as stroke.

Therefore, Twitter surveillance is a promising source for epidemic prediction and detection because of its sensitivity, transparency, and availability. In the future it can be integrated with epidemic intelligence services of public health authorities. However, nowadays it can easily produce exaggerated and misleading reports.

2.3.4. Twitter use in emergencies

Twitter is changing the way people communicate not only in their daily lives, but also during emergencies, including public health disasters. For example, during the Great East Japan Earthquake in 2011 Twitter was used for interpersonal communication, local communication (local governments, organisations, and media), for mass media communication, information sharing and gathering, and as a channel for communication between individuals, mass media, government, and the public. (80)

Twitter might be the fastest way for the information to reach the public and to build situational awareness at the time when a disaster occurs. Thus, during a Mississippi tornado official authorities immediately informed the public and urged them to seek shelter. (81) Twitter was also used by New York officials during Hurricane Sandy in 2012, and it was demonstrated how tweets were widely spread beyond existing follower bases. (82) Tweets were used to demonstrate how Twitter might help people to determine which medical facility is able to offer them needed medical help. For example, during the 2009 H1N1 outbreak one of the US health departments immediately tweeted where vaccination was provided and people went to these sites. (3) However, not all the information shared on Twitter about a disaster comes from official sources and some may not be accurate and trustable.

Geo-location options on social media tools may enable tracking of the actual location of people during the disasters. Using crowd-sourced information it is possible to “crowdmap”, that is to create real-time maps as incidents are occurring. (83) For example, in 2010 in Portland, USA, during the pipe explosion, the crowdmap, which was created using Google maps and tweets, helped police officers to identify the area of the accident. (83) Nevertheless, in the event of disasters, communication infrastructure may collapse and responders will still have to rely on radio-based communication technology.

Social media data may also provide many other insights during the emergencies. For example, Splunk4Good programme analysed tweets in order to understand how public fear changed as Hurricane Sandy in 2012 approached, the number of people asking for help, and the rate of people evacuating the area over time. (84)

In summary, Twitter is an important communication tool for public health organisations to rapidly communicate with a large part of their audience. They use it both in everyday work and in public health emergencies, including disease outbreaks.

2.4. West African Ebola virus disease outbreak

Ebola virus disease (EVD) is a severe, often fatal illness in humans caused by Ebola viruses, which develops a haemorrhagic fever in humans. (85) It is endemic in regions of central Africa. Small number of cases of haemorrhagic fever were described in 1976 from the outbreaks in two neighbouring regions: Southern Sudan and Northern Zaire (now Democratic Republic of the Congo) when Ebola virus was firstly isolated. (85) The latter occurred in a small village on the Ebola river, after which the virus takes its name. (86) The virus is transmitted to people from wild animals, and is spread in human population through human-to-human transmission. (86) The estimated average case fatality rate of the disease is 50%. (86)

The West African Ebola outbreak had been the most complex and the largest since the discovery of Ebola virus in 1976 by September 2015. Unlike the previous outbreaks, which affected remote villages in Central Africa, this one had affected major urban as well as rural areas. It was identified that the first case of the West African Ebola outbreak dated back to 26 December 2013, when a 2-year-old boy in a remote Guinean village, located in a triangle-shaped area where the borders of Guinea, Sierra Leone and Liberia converge, fell ill. (87) These three countries were deeply impoverished and their health infrastructures damaged due to years of civil war. After the first case, the disease remained undetected,

resulting in several chains of the virus transmission. The first cases were officially notified in March 2014 in Guinea, after which the disease spread across land borders to Sierra Leone and Liberia, by air to Nigeria and the USA and European Union countries, by land to Senegal and Mali. (86) The most severely affected countries of the West African Ebola outbreak had been Guinea, Sierra Leone, and Liberia. All of them had emerged from long periods of conflict and instability by that time, and had weak health systems and lack of human and infrastructural resources. (88) On 8 August 2014, the West African Ebola outbreak was declared a Public Health Emergency of International Concern¹² by the Director-General of the WHO. (80) Up to 20 September 2015, there were a total of 28,331 reported confirmed, probable and suspected cases of EVD in Guinea, Liberia, Sierra Leone, Nigeria, Mali, Senegal, Italy, USA, UK, and Spain with 11,310 reported deaths (Figure 11). (89) In the most affected countries, Guinea, Sierra Leone, and Liberia, a total of 28,295 cases had been reported with 11,295 deaths. (89) Whereas Liberia was declared free of Ebola virus transmission on 3 September 2015, (90) at the end of September 2015 the outbreak still remained in two countries: Sierra Leone and Guinea. The key events of the West African Ebola outbreak are represented in Table 1.

Although there was yet no licensed treatment proven to neutralise the Ebola virus, early rehydration and symptomatic therapy increase survival. On 31 July 2015 the WHO announced that the world is on the verge of an effective Ebola vaccine due to the promising interim results of the Phase III efficacy trial of one of the vaccines. (91)

¹² Public Health Emergency of International Concern is an extraordinary event that constitute a public health risk to other countries through international spread of disease and potentially require a coordinated international response.

Figure 11. Confirmed, probable and suspected EVD cases worldwide. (92)

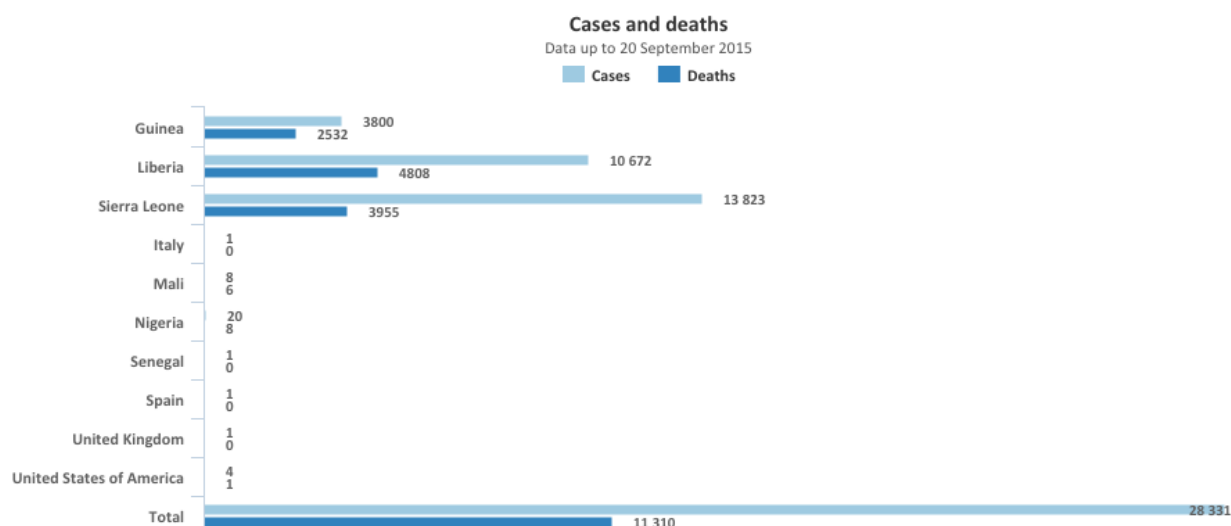


Table 1. Key events of the West African Ebola outbreak. (93)

Date	Event
26 December 2013	A 2-year-old boy in a remote Guinean village fell ill with Ebola. He was considered to be the index case in the West African Ebola outbreak.
19 March 2014	Guinean health officials announced the outbreak of a mysterious haemorrhagic fever. 35 cases were reported, and at least 23 people died.
22 March 2014	Guinean health official received a confirmation that a mysterious haemorrhagic fever was Ebola.
23 March 2014	The WHO released its first report concerning Ebola in West Africa.
31 March 2014	The first cases of EVD were confirmed in Liberia.
12 May 2014	EVD cases were reported in Conakry, the capital of Guinea and a city with a population of around two million people.
26 May 2014	WHO reported the first cases and deaths in Sierra Leone.
17 June 2014	Liberia reported that EVD had reached its capital, Monrovia.
20 June 2014	Doctors Without Borders declared the second wave of the outbreak "totally out of control".
24 July 2014	A Liberian man was being tested for Ebola in Lagos, Nigeria.
25 July 2014	Nigeria reported its first Ebola fatality. Moreover, the first case in Freetown, Sierra Leone's capital, was recorded.
1 August 2014	It was announced that two Americans infected with Ebola in Liberia are being medically evacuated to Atlanta, Georgia for treatment at Emory University Hospital.
2 August 2014	First infected with EVD American missionary arrived to the US.
5 August 2014	Second infected with EVD American arrived to the US.
8 August 2014	WHO declared the EVD epidemic a Public Health Emergency of International Concern.
12 August 2014	A Catholic priest who had been medically evacuated from Liberia where he had been volunteering died in Madrid.
12 August 2014	The WHO permitted the use of experimental drugs to treat Ebola.
21 August 2014	The two first medically evacuated cases in the US, having been successfully treated with the experimental therapy ZMapp, were released from Emory

	University Hospital free of the EVD.
24 August 2014	A British man was evacuated from Sierra Leone to an isolation unit in the UK.
27 August 2014	A Senegalese epidemiologist who was working in Sierra Leone for the WHO was transferred to an isolation ward in Germany.
29 August 2014	Senegal confirmed its first EVD case.
16 September 2014	USA President Barack Obama announced that the US would send 3,000 troops to West Africa to build Ebola treatment centres and establish a military coordination centre.
19 September 2014	The first French national to be infected with EVD was medically evacuated from Liberia to Paris.
22 September 2014	A Swiss health worker was flown out from Sierra Leone to the University Hospital of Geneva after being bitten by an Ebola-infected child.
24 September 2014	The first EVD case in the US visited the emergency room in Dallas, where he was diagnosed with a “low-grade, common viral disease” and sent home with antibiotics.
30 September 2014	The CDC announced the first case of EVD in the Americas, in Dallas, Texas.
6 October 2014	MSF announced that one of their workers was to be transported to Norway for treatment.
6 October 2014	An auxiliary nurse, who had cared for a patient in Spain, tested positive for Ebola, marking the first transmission of EVD outside of West Africa.
8 October 2014	Man diagnosed with Ebola in Dallas, Texas died of the disease.
10 October 2014	CDC announced that a healthcare worker who provided care for the index US patient tested positive for Ebola.
15 October 2014	A second healthcare worker who provided care for the index US patient tested positive for Ebola.
15 October 2014	The news was spread that a second Ebola-infected health worker in Dallas took a flight a day before she was diagnosed with Ebola.
20 October 2014	WHO declared Nigeria free of Ebola virus transmission.
23 October 2014	Ebola case was reported in a medical aid worker who had returned to New York City from Guinea.
10 December 2014	Time magazine names The Ebola Fighters its Person of the Year.
29 December 2015	Healthcare worker was first to be diagnosed with Ebola in the UK.
10 March 2015	WHO declared UK free of Ebola virus transmission.
13 March 2015	Healthcare worker who contracted Ebola in Sierra Leone arrived to Maryland hospital, USA
9 May 2015	WHO declared Liberia free of Ebola virus transmission.
29 June 2015	Ebola was back in Liberia: 1 month, 20 days after “free” declaration.
28 July 2015	Liberia began another countdown to being Ebola virus transmission free.
31 July 2015	Ebola vaccine trial proved 100% successful in Guinea.
2 August 2015	The number of new Ebola diagnoses in Sierra Leone and Guinea reached its lowest point in over a year.
3 September 2015	Liberia declared free of Ebola virus transmission.

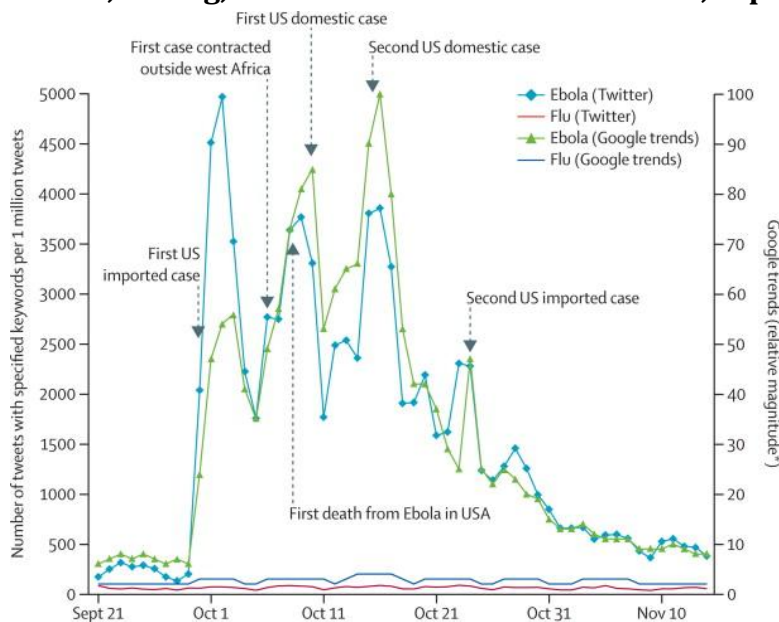
2.5. Ebola on Twitter

Despite the level of poverty in the countries most affected by the Ebola outbreak, the use of tablets, computers, and smartphones in these regions is increasing, which creates an opportunity to communicate public health information on social media. (10) For example, Nigeria, the most populous country in Africa with the population of over 174 million people, has over 114 million mobile phone subscriptions. Almost 56 million Nigerians are regular Internet users mostly through mobile devices, with a growing interest in social media. (94) In 2014, Nigeria accounted for approximately 1.8 million Twitter users. (95) The control of the Ebola outbreak requires effective communication strategy and this is where, many believe, social media may play a significant role. (96)

The Ebola outbreak largely failed to receive attention from the Western press, the Western government websites, and social media until late July 2014 when two American health workers contracted Ebola and the “outbreak” suddenly started to be called an “epidemic” in global headlines. (97) Initiated by media attention, fear of Ebola started to spread globally faster than the virus. The global Twitter conversation about Ebola “exploded” on 30 September 2014, when a Liberian man who recently travelled to the USA from Liberia was diagnosed with Ebola. (98) On the night of 1 October 2014 the conversation reached the rate of 6,000 tweets per minute (in comparison with 100 per minute before 30 September 2014). (99) Between 7 October and 3 November 2014 Ebola outbreak received an extensive and continuous coverage in the US media with nearly 1,000 segments about Ebola aired. (100) During October 2014 large tweet frequency peaks were observed in Ebola conversation. Figure 12 demonstrates how during October 2014, Twitter and Google traffic about Ebola increased, reacting to the news about each of the four Ebola cases in the USA (in comparison with influenza). (5) Social media amplified fear about the imported Ebola case. (5) The coverage of the Ebola outbreak on the news was voluminous, but far from balanced and often narrow. The cases of Ebola that were brought or initially

appeared in Europe or the USA received much more attention than the thousands of cases in West African countries. (101) Even reputable news organisations of the USA and the UK rarely supported their articles with science, and contributed to public confusion and misinformation about the risk of exposure and the real threat of the disease. (101) Although the risk for most Americans was low, the public expressed anxiety. Each Ebola-related news video inspired tens of thousands of Ebola tweets and Internet searches. (102) This coverage was criticised by some US journalists who highlighted how dramatically disproportionate it was to the number of people it affected in the US. (100)

Figure 12. Temporal trends on Twitter and Google about Ebola and influenza (flu) before, during, and after Ebola cases in the USA, September to November 2014. (103)



One of the reasons for such high influence of the Western news on the Twitter conversation is the fact that Twitter and the Internet have much higher penetration in the developed world, including Europe and the USA, than in Western African countries, especially their rural areas. (104) Therefore, it is not a surprise that those who were most affected in the West African Ebola outbreak remained silent on social media, and many of the messages disseminated on Twitter did not reach them. (105)

Interestingly, it has also been shown that there was a case when Ebola news on Twitter was shared even before the official announcement from governmental authorities.

Thus, the first probable EVD case was officially announced in Nigeria on 25 July 2014, whereas Ebola tweets started to rise in Nigeria on 24 July 2014 when the first news about EVD in Nigeria was tweeted. (9)

However, it has been demonstrated that Twitter can still also be used as a tool for Ebola surveillance. In the first couple of months, before Ebola was spread to the US (which generated a lot of background noise of Twitter), it was possible to predict the spread of the Ebola virus in the affected countries (based on the comparison of the Twitter, WHO and CDC data). Moreover, it was suggested that with the reduction of the background noise over time the accuracy of predictions would increase again. (106)

Since the beginning of the Ebola Twitter conversation it had been noticed that a lot of misinformation about Ebola was being spread on this platform. Among the tweets from Guinea, Liberia, and Nigeria containing the words “ebola”, “cure”, and “prevention” in the first week of September 2014, more than half contained misinformation. (10) Among the most common misinformation were the messages that Ebola can be treated by ewedu¹³ plant or by blood transfusion, and by drinking and washing in salty water. The latter rumour even caused two deaths at the beginning of August 2014 in Nigeria, as reported by the WHO. (107) The WHO also acknowledged that “the scale, duration and lethality of the Ebola outbreak have generated a high level of public fear and anxiety, which extends well beyond West Africa”. (107) Moreover, on 14 August 2014, the US Food and Drug Administration also warned the public that there were no approved treatments or vaccines against Ebola, saying that they received a number of consumer complaints about a variety of products claiming to either treat or prevent the disease. (108) However, a study focused on Ebola tweets from October 2014 suggested that there were 4.5 times more tweets containing credible information on Ebola versus tweets that were regarded as speculative.

¹³ Ewedu is a flowering herb; its leaves are eaten throughout West African countries in soups and sauces.

(6) Nevertheless, these studies used different methods of identifying misinformation, which might make their findings incomparable.

In a recent survey-based study (109) it was demonstrated that most Americans used the Internet (96%) and media (86%) as major sources of new information about Ebola in 2014. The Internet was also rated as the most trustworthy source of Ebola information. This fact provides opportunity for trusted sources, such as WHO and CDC, to spread scientifically accurate and reliable information to the public through their preferred channel. Therefore, it will be in the primary interest of the global health organisations such as the WHO and CDC to correct any rumours and misinformation on the web as soon as possible.

While some of the tweets were left undisputed, others were corrected by the government agencies, such as a Nigerian government agency. (10) Civil society and social media organisations started to take action as well. For example, a Nigerian dentist created campaign with a hashtag @EbolaAlert in order to disseminate accurate information about EVD on Twitter. (4) The campaign then was spread to other social media platforms, like Facebook. Facebook also provided a button, which enabled people around the world to donate money to fight Ebola. (4) In order to maximise their efforts, WHO and CDC provided accurate information about Ebola on their social media channels in addition to their traditional communication channels. In collaboration with their partners (local, national, international) they shared and amplified key messages through social media. For example, they provided information to the Ministry of Health of Nigeria, local healthcare organisations, popular bloggers, and others with large numbers of followers on Facebook and Twitter. (4)

CDC has a research team monitoring trends on social media seven days a week to identify rumours, misinformation, gaps in information and emerging topics to inform their overall communication strategy. (110) Since the beginning of August 2014, CDC had

disseminated over 5,600 social media messages across multiple channels. 200 of these messages, containing visuals and links to resources, were during the first week of August 2014 when the first Ebola cases were evacuated to the USA from West Africa. When misinformation was identified, response-messages were developed in collaboration with the CDC content experts. One piece of such misinformation questioned why healthcare workers wear so much personal protective equipment if Ebola is not an airborne disease. Public health messages about the necessity of wearing this equipment due to the large amounts of blood, vomit, and other bodily fluids in the environment surrounding health workers were disseminated. (110) CDC organised four Twitter chats¹⁴, three targeting the general population, and one targeting health care providers, using the hashtag #CDCchat. The potential reach of these chats was 109 million (4 August 2014), 37 million (8 August 2014), 161 million (2 October 2014), and 100 million (8 October 2014) participants. (110) The chat on 2 October 2014 was organised in order to answer the questions from the general public about the first Ebola case diagnosed in the US. It included 7,484 participants, and CDC answered 155 questions. (110) Not only did these chats enable the dissemination of the necessary information to the public, they also enabled the organisation to engage with their audience and hear their questions, concerns, and opinions.

However, the current capability of global organisations to spread their information on Twitter might be limited. For example, a CDC infographic tweet that was posted on 30 September 2014 explaining how EVD is spread was retweeted more than 4,000 times. Interestingly, however, a humour-focused Twitter account “Tweet like a girl” retweeted a CDC “Facts About Ebola” image and was retweeted more than 12,000 times, more than the original tweet posted through the CDC account. (8)

The WHO organised a variety of Ebola social media campaigns on several social media platforms, providing factual information about the disease to the public around the

¹⁴ A Twitter chat is a Twitter discussion that takes place in real time at a pre-arranged time on a pre-arranged subject. (111)

world and updating them about the current situation. In September 2014, WHO and CDC were among the top 10 accounts on Twitter most often mentioned among the messages about Ebola. (7) However, the WHO social media campaigns on Ebola have not been described in the literature.

2.6. Study rationale

Overall, existing evidence demonstrates that Twitter has played a role in the spread of information during the West African Ebola outbreak. Twitter conversation about Ebola was initiated and amplified by Western media coverage. In October 2014, when the first Ebola case was diagnosed in the US, the Twitter conversation “exploded”. Governmental health organisations responding to the outbreak at the global scale, such as the WHO and CDC also used Twitter as their official communication tool. However, researchers have only analysed either short periods of the conversation, or incomplete Twitter data. Moreover, the search strategies, data filtering and cleaning methods have been unclear. The literature on Twitter engagement related to the West African Ebola outbreak and the ways in which global health organisations used the platform in this period is also lacking.

The rationale behind the present study is to assess the ways in which Twitter was used during the West African Ebola outbreak by the general Twitter population, as well as by the WHO and CDC in order to better inform future efforts to harness the public health potential of Twitter. This study will also aim to identify the geography of Ebola tweets, the accounts that demonstrated the highest level of engagement on Twitter related to the West African Ebola outbreak, and the place of the global health organisations among them. In particular, it aims to investigate the level of engagement achieved by the WHO and CDC in relation to Ebola Twitter conversation. Both WHO and CDC are of interest for this study, as these governmental organisations participated in Ebola response at the global scale and they use Twitter as one of their official communication tools. CDC, being a US health

agency, also managed the Ebola cases and public anxiety in the US, the country whose citizens seemed to have the highest contribution to the Twitter conversation about Ebola.

Chapter 3. Methodological challenges of accessing and analysing Twitter data

In order to answer my RQs, I performed an overview of methodological challenges of accessing and analysing Twitter data that can be applied to my project.

3.1. Data sources, attributes, and costs

Twitter data can be obtained from different sources, which influence the costs of its cleaning and processing and the representativeness of a tweet sample. It can be retrieved using the Twitter search option and Twitter search application programming interface (API), Twitter streaming API, Twitter data resellers, and automated dashboard vendors.

3.1.1. Twitter search option, Twitter search API, and Twitter Streaming API

The Twitter data can be retrieved manually or automatically using keywords or usernames through the Twitter Search option (twitter.com/search), Twitter search API or Twitter streaming API¹⁵. Twitter search and Twitter search API options allow requests for recent or popular tweets and provide those tweets that have already occurred, and are focused on relevance, not completeness. Streaming API, unlike search API, gives researchers the access to the global stream of Twitter data and happens in real-time. Tweets are pushed directly to a user if they match a set of registered criteria (keywords, username and locations).

Although the strengths of these methods are their simplicity and low cost, the major limitation of these options is that they do not deliver all the tweets that meet the searching criteria, (112,113) and documentation from Twitter on what and how much data users can get is lacking. Therefore, findings of studies which use this method of data retrieval cannot be generalised to the overall Twitter population (114) and they cannot be used to estimate

¹⁵ API is an instruction set which was created for developers to interact with technology.

the absolute values (for example, the most retweeted tweets, the number of tweets posted by an account of interest). It is also almost impossible to use this method to collect historic data over a long period of time (months and years).

3.1.2. Commercial vendors

Twitter licenses its full data stream to commercial vendors that provide clean Twitter data and basic automated analysis in real time via automated dashboards. Examples of such vendors are Topsy¹⁶, DemographicsPro¹⁷, Hashtagify¹⁸, and Symplur¹⁹.

The advantage of using vendor services is that they provide easy-to-use dashboards with visualisation tools, such as word clouds, figures, and graphs, which analyse the data in real time. (115) In addition to tweet metadata, it is possible to perform an analysis against augmented data. For example, DemographicsPro augments tweets with the demographic information, relying on the “ability to estimate or infer the likely demographic characteristics (gender, age, marital status, ethnicity, occupation, and location) of local consumers based purely on their social media presence/usage”. (116)

These platforms limit the access to tweet data and to the filtering criteria and it is only possible to perform analysis using dashboards or algorithms developed by them without seeing the raw data. Several studies used Topsy platform for their analysis (117–119), with which they could observe the frequency of tweets over the desired time period, and performed sentiment analysis. However, they were able to use only provided dashboards and algorithms developed by Topsy, which could not be verified. A study that focused on characterisation of pro-marijuana Twitter account followers used DemographicPro platform in order to predict demographic characteristics of followers of

¹⁶ <http://topsy.com/>

¹⁷ <http://www.demographicspro.com/>

¹⁸ <http://hashtagify.me/>

¹⁹ <http://www.symplur.com/>

that account. (120) However, these characteristics were not reported by the followers, but inferred based on their social media use and cannot be assumed to be accurate.

3.1.3. Twitter data resellers

Gnip²⁰ and DataSift²¹ were the leading enterprise social data platforms that used to resell Twitter firehose²² data until August 2015, when Twitter ended its partnership with DataSift. At the time of the data collection for this study, these platforms offered many of the same social data sources and provided access to the full historic and real-time Twitter firehose.

Undoubtedly, the biggest strength of these data sources is that they provide access to the full Twitter dataset, giving researchers control in retrieving, storing, and analysing the data. Filters offered by the platforms help to extract the data sample researchers are interested in and clean it. Augmentations enrich the data, allowing researchers to include tweets more relevant to the question of interest. Using such platforms, it is possible to get initial insights on the data before running the actual queries through the “historic preview” service. However, this is the most expensive source of Twitter data, the cost of which is determined primarily by Twitter itself. A lot of resources also may be required to build a system to store and analyse the data. Furthermore, it still remains possible that not all of tweets will be delivered by the Twitter interface through these platforms, and this is impossible to verify. In addition, researchers have to trust blindly to the algorithms behind the filters provided and, therefore, the results delivered.

To sum up, unlike other data sources, data resellers enable acquisition of full data samples, providing full tweet metadata and enriched data as well as other services enabling data filtering and cleaning. Researchers are able to fully control and manage the data. The

²⁰ <https://gnip.com/>

²¹ <https://datasift.com/>

²² A firehose is all of Twitter's public tweets.

platforms also serve as the best option for obtaining historical data over a long period of time. Although their main limitation is their relatively high cost, it was offset by the generous purchase of the data by Spunk4Good²³ through their existing DataSift account. Therefore, I used DataSift as a data source for this study.

3.2. Sampling considerations and filtering

In order to identify relevant data, search terms and hashtags that reflect common terms for a research question need to be considered. Additional online tools, such as Hashtagify.me might be used to identify relevant hashtags. (121) The list of search terms must not over- or underestimate the volume of discussion about the topic. Basic qualitative review of the retrieved content can help to determine whether the used search terms are appropriate. Language patterns, such as slang, common abbreviations, and sarcasm are also important to consider. For these purposes Urbandictionary²⁴ can be used to identify relevant keywords. (122)

Not all the tweets containing selected keywords or hashtags might be relevant for the research question, so data filtering criteria (inclusion and exclusion criteria) must be applied. For example, it is possible to filter tweets by language, date, geo-location, type of tweets (original tweets or retweets), user, and other criteria based on tweets' metadata. Whereas filtering can be done manually for a small sample of tweets, computational techniques might be applied for larger samples of tweets. DataSift, which is used for this project, provides computational services for data filtering, which can be performed both against tweet metadata and augmented data.

Search query can generate a large volume of data. In some cases, full samples can be analysed in a research project, whereas in other cases, smaller samples might be required for a study. Different methods of sampling have been investigated: a random sample of all

²³ <http://splunk4good.com/>

²⁴ <http://www.urbandictionary.com/>

tweets, probability proportionate to size based on number of followers, a random sample of most followed Twitter users, and a random sample of only original tweets after retweets removal. The first two demonstrated to be the most representative of the full Twitter sample. (123) I used a sample of all tweets for this study.

3.3. Data management and cleaning

One of the major challenges for Twitter analysis is the management and cleaning of a large volume of data. Some of the datasets may include millions or even billions of tweets and many traditional software packages cannot handle this number of cases. Data storage is a challenge. Whereas, software that can manage large amounts of data exist (statistical software R, text analysis with Python, Apache Hadoop for data processing, and Splunk Enterprise), this may require a lot of human, technological and financial resources to build the environment necessary for data analysis.

Building a quality Twitter dataset also requires a significant amount of resources. For example, for content-mining tools it is important to remove all nonstandard characteristics or special characters. (124) Twitter also has noisy linguistic characteristics, as it has a variety of languages, official and unofficial abbreviations, and unstructured grammar. The other source of noise on Twitter is spam. Although, Twitter implements strategies to limit spamming accounts, it cannot fully prevent it. Therefore, researchers try to implement algorithms to clean the data of spam. (125,126)

Splunk Enterprise 6.3.0, the Platform for Machine Data, a time-series optimised NoSQL²⁵ data storage, analytics and visualisation software, was chosen for the project as software for data management and analysis. It has the rare capability to interpret a

²⁵ "NoSQL is a class of database management systems (DBMS) that do not follow all of the rules of a relational DBMS and cannot use traditional SQL to query data. The term is somewhat misleading when interpreted as "No SQL," and most translate it as "Not Only SQL," as this type of database is not generally a replacement but, rather, a complementary addition to RDBMSs and SQL". (4) "Structured Query Language (SQL) is a standard computer language for relational database management and data manipulation. SQL is used to query, insert, update and modify data". (5)

"Schema-on-the-fly"²⁶ without pre-parsing or any ETL²⁷ processing. This enables a user to easily manage the data and focus on the work with it. Although the platform enables the analysis of really large datasets in a manageable and quick way, it requires technological resources in order to prepare the dataset for the analysis.

3.4. Data analysis

Both qualitative and quantitative types of analysis can be applied to Twitter data. They will be discussed in this section.

3.4.1. Quantitative methods of Twitter data analysis

Descriptive statistics

Tweet metadata (Figure 5) and augmented data can be used to perform descriptive analysis considering various metrics of interest. For example, it is possible to estimate user metrics, such as the number of tweeting accounts about a given topic, number of followers of tweeting accounts, number of messages sent by a user, and types of tweets sent.

Temporal metrics, such as the number of tweets sent in a particular period, can be investigated and activity peaks can be analysed to identify events leading to the increase. This was done in a study that investigated which real-world events make people talk about antibiotics, (119) health reforms, (59) or about Ebola. (79) As has been previously discussed, conversations on Twitter can be influenced by media coverage. Depending on the event, it can be appropriate to estimate temporal metrics minute-by-minute (live sports, television shows), day-by-day (election campaigns, natural disasters), or month-by-month (military conflicts, policy implementation, outbreaks). (114) It is possible to examine the total number of tweets sent in different periods, unique users active per

²⁶ This means without a prior knowledge of a schema, Splunk can infer one if possible.

²⁷ "ETL is short for extract, transform, load, three database functions that are combined into one tool to pull data out of one database and place it into another database". (127)

period of time, or tweets sent by each user during a particular period. (114) However, these short study periods may not fully represent daily variations in public opinion and that Twitter use in different time zones likely peaks at different times of a day and a week. (58) When performing a longer-term analysis, it is important to adjust for the growth in Twitter users over time. This can be done based on published Twitter statistics. This helps to prevent bias and makes the Twitter study results comparable to the results of the other data source analysis. (115) For example, after adjustment for the volume of tweets about flu over time, the trends of Twitter data volume more accurately matched the patterns of Google search about flu. (128)

Twitter activity in different geo-locations can be analysed or a map of spatial distribution of tweets can be generated. For example, this has been done in a study which mapped tweets about obesity across the USA. (129) The study demonstrated that higher densities of tweets were located around bigger cities on Eastern and Western coasts of the US. It has been suggested that geographical location of the tweets helps health scientists to better monitor and understand health status in certain locations. (129) Even though less than 3% of tweets are geo-located, (130) there are computational techniques that have been developed to provide further location information. (131)

These descriptive statistics are applied in this study in order to answer the RQ1 and investigate the pattern of Twitter use in the general Twitter population.

URLs

Tweets may include URLs, which may provide reference to, for example, a webpage, a picture, or a video. URLs in tweets are usually shortened by various “URL shorteners²⁸”. Each of them gives URLs different unique identifiers²⁹. Data vendors provide services that may resolve all shortened links to an end point, converting them back to their original

²⁸ Services enabling shorten URLs in length, still directing to a required page.

²⁹ Multiple tweets referencing the same URL will link to different URLs.

URLs. (36) In a study of a large Twitter dataset over 7 years it was demonstrated that the prevalence of tweets containing URLs had decreased since 2009, when approximately 24% of tweets contained at least one URL, and stabilised at approximately 12% in January 2014. (25) These findings were supported by the studies on other large Twitter datasets. For example, in 2010, 21.1% of 73 million tweets contained at least one URL, whereas 28.4% of retweets did. (132) The fact that the prevalence of URLs among retweets is higher than that among the tweets suggests that tweets are more likely to be retweeted if they contain a URL, and that retweets are used to spread interesting web pages, videos and other web content to Twitter users and may stimulate Twitter engagement. In 2012 it was shown that 15.9% of tweets contained URLs. (36)

The fact that a substantial proportion of tweets contain URLs to external web pages, which often offer larger content, might be used to explore other research questions. For example, it has been demonstrated that URLs might be used to study geographical affinity in terms of topics users discuss, by looking at the location of users and geographical focus of their tweets identified through URL instead of short tweet texts. (36)

Twitter engagement measurements

Different ways exist in order to measure Twitter engagement. Whereas it is possible to count the number of tweets and the number of Twitter user followers and followees in order to estimate the degree of their engagement with other users, it would not be fair to assume that users pay equal attention to the tweets from different accounts. Looking at the relative performance of Twitter users based on the number of tweets they post also may not be possible due to an increasing number of accounts of automated tweet aggregators. According to the Twitter statement in August 2014, up to 8.5% (23 million) of all active users used third party applications that may have automatically contacted Twitter servers and aggregated Twitter content “with no user action involved”. (133) This kind of account may constantly post vast amount of tweets and distort the results. Increase of malicious

behaviour on Twitter since 2009 with the existence of so called Twitter account markets that produce fake followers and fake accounts makes it harder to use these numbers as a measure of Twitter account popularity or engagement. (25,134) It has been proposed, that, rather than simply the number of followers someone has, Twitter engagement might be better estimated by looking at the active interactions via usage of different communicative tools available to Twitter users, such as original tweets, retweets, and replies. (135)

For the information dissemination the mechanism of retweeting was firstly created by users and officially adopted by Twitter in 2009. (25) Retweets are seen by a retweeter's followers. There might be different motivations for retweeting: amplification or spread of tweets to new audience, entertainment or informing a specific audience, commenting on someone's tweet, public agreement with someone, act of friendship, loyalty, self-gain (gaining followers), saving tweets for future personal access and others. (136) What people retweet also varies and is usually linked to the reason why they retweet. For example, a standard response to the emergence of breaking news or other news events is the tendency to find, share, and "reshare" relevant information, resulting in a high prevalence of retweets and URLs in them. (136) Furthermore, many retweets appear in order to encourage social action, for example, to protest, to donate, or as a social media campaign in order to be heard. (136) Moreover, tweets containing photos, videos, quotes, digests, and hashtags receive more retweets. For example, tweets posted by governmental or political verified accounts are 62% more likely to be retweeted if they contained a photo URL. (137)

Twitter users may start or join Twitter conversation by posting a reply to other users or by mentioning them. "A mention is a tweet that contains another user's @username anywhere in the body of the tweet". (138) Mentions are subsequently translated into clickable hyperlinks to the mentioned users. Mentions are seen by mentioned users and by the followers of the sender. They can be searched in a search box by a "@username". A reply is a response to another Twitter user by starting a message with a "@username" of a

person the reply is sent to. Twitter provides an official reply service. (138) The replies are seen in the timelines of the sender, the recipient, and the followers of both of them. They are also seen on the page of a sender in “tweets & replies” section. In 2010, on a sample of 1.2 billion tweets it was demonstrated that 92.4% of retweets and 96.9% of replies happen within the first hour of the original tweet being published. (139) In a study in 2014 on a smaller sample of tweets it was also demonstrated that more than a half of retweeted tweets get more than 50% retweets within one hour after their creation. (140)

In summary, the number of retweets might act as a measure of tweet popularity, and as Twitter user popularity and their level of engagement with other Twitter users. The number of mentions and replies received by a Twitter user might measure the level of direct engagement with other Twitter users.

The prevalence of mentions, retweets, and replies changes over time. In a study of the evolution of Twitter users and their behaviour on a set of 37 billion tweets over 7 years (January 2006 to January 2014) it was demonstrated that the prevalence of tweets with “any mentions” (including retweets, replies, and all other mentions) (Figure 13) and the prevalence of retweets (Figure 14) had increased substantially over the study period. (25) As of January 2014, 50% of tweets contained “any mentions” and 27% of tweets were retweets, whereas the percentage of tweets containing URLs had decreased and was approximately 12% in January 2014. (25) This suggests that Twitter had become more conversational. However, as demonstrated in Figure 14, the prevalence of replies (24% in January 2014) had been decreasing, suggesting that person-to-person communication had been declining. (25) Although the samples of tweets for this study came from different sources and were not random and full, their findings were supported by other studies. For example, in 2012, in a sample of 1.5 billion tweets 55.1% of them contained “any mentions”, and more than 23.9% were retweets. (36) 12.4% of users active during this period were mentioned, and 4.5% of users were retweeted. (36)

Figure 13. The percentage of tweets with different types of entities over time. (141)

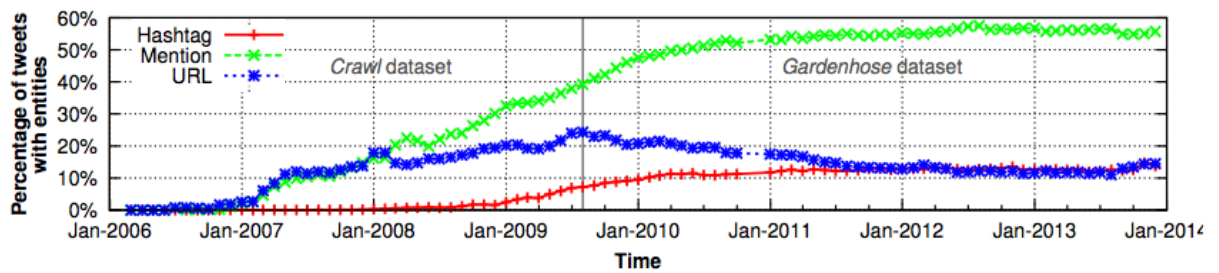
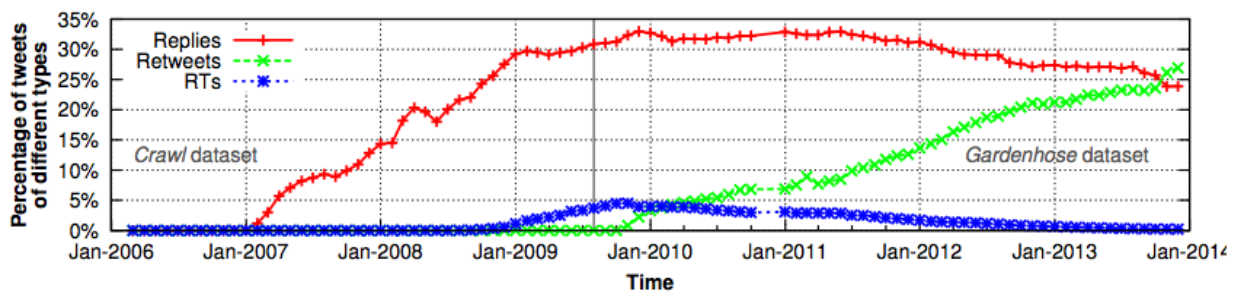


Figure 14. The percentage of tweets of different types over time. (142)



Both RTs and Retweets represent retweets, with the former capturing manually created retweets by users. Native retweets were supported by Twitter starting in November 2009.

Through analysis of thematic Twitter samples from the same time period, it was demonstrated that engagement parameters may also differ due to different thematic and contextual factors the Twitter conversation occurs in. (136)

For this study, I decided to investigate mentions, replies, and retweets in order to answer the RQ2 and RQ3 and estimate Twitter engagement of the general Twitter public, identify the accounts with the highest level of engagement, and study the engagement of the general Twitter population with the accounts of the WHO and CDC over the study period.

3.4.2. Qualitative analysis of Twitter data

Account description and the tweet content can be analysed qualitatively. From an account description it might be possible to draw a conclusion about the account type (an individual or an institution). This classification might be expanded further, identifying particular categories of individuals (representatives of civil society or key players in the area of interest) or institutions (governmental, non-governmental, private sector, or

media). This information can be used to compare the relative performance of institutional or individual accounts in a variety of communicative contexts on Twitter. For example, in a study that analysed Twitter activity during breast cancer awareness month, Twitter users were classified into three categories, individuals, organisations, and celebrities, in order to determine which of them was the most influential during the campaign. (143) Results showed that celebrities reached more people in comparison with other groups despite the fact that they posted fewer tweets. (143)

The content of a tweet itself is usually of primary interest for researchers. Content analysis of tweets has been performed in a large number of studies analysing health questions on Twitter. (71) Qualitative content analysis is a method for systematically describing the meaning of qualitative data. This is done by assigning successive parts of the material to the categories of a coding frame. This frame is at the heart of the method, and it contains all those aspects that feature in the description and interpretation of the material. (146)

Performance of these analyses on tweets is challenging due to their brevity and the presence of unstructured grammar, slang, sarcasm and other unconventional forms of written expression. (115) Although automated analysis with the use of computational techniques (natural language processing, machine learning, predictive analytics) can be used for the analysis of large datasets, it requires computational linguistic expertise. It is also possible to purchase the services of commercial vendors for the content analysis of large datasets, but existing computational algorithms have been mostly optimised to code content about commercial products and provide least accurate results for health research. (115) Manual coding of tweets is ideal for the analysis of small data samples as it is most accurate (especially if it is done by two trained coders), but it requires time, human, and financial resources.

Content analysis is used in this study in order to answer the RQ2 and RQ3, and to categorise the accounts that demonstrated the highest level of engagement and understand how Twitter was used by the WHO and CDC during the West African Ebola outbreak.

Chapter 4. Methods

4.1. Data source

4.1.1. DataSift

Through the Splunk4Good corporate social responsibility programme, Splunk Inc. donated enriched Twitter Social data for this study. These data were purchased from Twitter data reseller platform DataSift (MediaSift, Inc) that provides access to the social data from over 20 of the most popular social networks and many other sources. From January 2010 until August 2015 (147) the company had access to the Twitter firehose³⁰ (148) via a direct connection to Twitter. DataSift filters and aggregates data from several sources. In addition to extracted metadata, it further provides various augmentations to enrich the data, which are described in the Appendix 2. In order to perform filtering on DataSift, the company developed their own curation language called Curated Stream Definition Language (CSDL). It allows users to create filters with unlimited precision and enrich the data as needed.

4.1.2. DataSift historic previews

I used the DataSift “Historic Preview” service as an initial step in this project. This service enabled me to perform a historical analysis targeting the fields of interest in data without returning the actual data points from the platform. It is possible to run a filter quickly and estimate the quantity and quality of the data on a 1% random sample before running the actual query. (149) For sampling, DataSift internally generates a floating-point random number between 0 and 100. Historic previews may analyse the data within a period of up to 32 days.

³⁰ The stream containing all the tweets that is not publically available.

Using CSDL, I ran historic previews within a period from 1 December 2013 to 12 May 2015 in order to define the filtering criteria and to get initial insights of interest for the project. I specifically examined tweet text, volume of tweets and retweets, Twitter user description, augmented demographics characteristics (location, age, and gender), top 10 Twitter accounts, hashtags, links, mentions, languages, and sources³¹ of tweets on a 1% random sample of the data. An example of the historic preview output is provided in the Appendix 3. Based on these results, I tested, modified, and defined the filtering criteria. One of the decisions that I made was to include tweets using links augmentation of the DataSift, which added approximately 7.5 million tweets to the dataset, which eventually comprised approximately 11.5% of the tweets in the final dataset.

4.2. Sampling considerations and filtering criteria

I used the following filtering criteria in order to identify the relevant data. Firstly, I included all the tweets containing the word “ebola” (Latin script) in the dataset. Secondly, I identified from the literature, media coverage and social data platforms such as Hashtagify.me³², the most relevant hashtags about EVD. I included the following hashtags: #stopebola, #ebolafacts, #ebolachat, #nursesfightebola, #ebolaethics, #ebolaresearch, #ebolanews, #fightebola, #ebolaresponse, #ebolavirus, #ebolawatch, #ebolaalert, #ebolaoutbreak, #ebola. Thirdly, I used the DataSift Links augmentation in order to include the tweets containing the URLs to the articles that had the word “ebola” in their title, description, or among their keywords, but not necessarily mentioning the word “ebola” or one of the hashtags in the tweet text.

³¹ The application that was used to generate and send a post.

³² <http://hashtagify.me/>

4.3. Data ingestion and deployment

In acquiring data, a stream of data in JSON format first went from Twitter directly to DataSift. DataSift performed data enrichment, adding augmented fields to the JSON object, and stored them in their archive. After the targeted queries were sent to Splunk team members and submitted by them in to the DataSift processing platform, DataSift pushed results directly in to Splunk, and they were delivered to Amazon's S3³³ (Simple Storage Service). For this study, the queries executed by DataSift were enriched by interaction, language, and links augmentations. The interaction augmentation enabled the grouping of the data points for tweets and retweets. The language augmentation analysed the tweet text in order to identify its language, as the language stated by a Twitter user (if any) is not always the actual language they tweet in. The links augmentation resolved any links within the content to their final endpoint. It also aggregated the content the links were pointing at so that the filtering could be performed against it. The sample data for this project spans from 1 December 2013, at 08:00:00 UTC³⁴ to 12 May 2015, until 06:00:00 UTC. The start date was determined by the month the West African Ebola outbreak began³⁵. The last day in the sample was determined by the date the data were obtained. The initial file contained 83,141,282 tweets across over 33,000 files (ranging 10 kB to 200 MB) and in total was about 250 GB in size. Splunk team members deployed cloud instance and storage in to Amazon Web Services³⁶, installed Splunk Enterprise 6.3.0 (further, Splunk), and ingested the data stored in Amazon's S3. Splunk is the platform for machine data, a time-series optimized NoSQL data storage, analytics and visualisation software. While receiving time

³³ <https://aws.amazon.com/s3/>

³⁴ Coordinated Universal Times

³⁵ The first case of EVD in Guinea was reported on 26 December 2013.

³⁶ <http://aws.amazon.com/>

series data (such as tweets) Splunk indexed³⁷ it, breaking it into events³⁸, based on the timestamps the tweets were created (the only identifiable field needed by Splunk for ingestion). (151) Splunk team members further assisted me in all setup, installation, configuration, data model development, training, and technical support.

4.4. Data analysis on Splunk

Event data³⁹ on the Splunk interface included tweet metadata and augmented fields. An example of the full event data output is presented in the Appendix 4. For the data analysis on Splunk, two types of searches can be performed: raw event searches and transforming searches. Raw event searches retrieve events from an index or indexes. Transforming searches perform some type of statistical calculation against a set of results and require fields and at least one of the statistical commands. Moreover, the searches can be sparse (looking for a single event that occurs in the dataset) or dense (which scan and report on many events). (153) The Search Processing Language (SPL) is used on Splunk for performing searches. It encompasses all the search commands and their functions, arguments, and clauses. (154) Data in Splunk are indexed chronologically, and it is therefore possible to select time ranges to apply to a search command. (155) Time field in tweets is based on the UTC time zone. Transforming searches return data in statistical tables required for charts and other types of data visualisation provided in the Splunk interface, such as time charts, charts, pie charts, and world maps. (156)

³⁷ The act of processing raw data, transforming it into events, and placing results into an index. An Index is “the repository for data in Splunk Enterprise”. (150)

³⁸ An event is a single piece of data in Splunk software. Each event consists of fields³⁸. When searches or queries are run in Splunk, they first start as a keyword search. In the process of the search, Splunk’s engine performs automatic field extraction where it can. This means without a prior knowledge of a schema, Splunk can infer one if possible.

³⁹ “A descriptive term for all the IT data that has been added to Splunk software indexes”. (152)

In order to manage the data on Splunk more easily and quickly I created the accelerated data model⁴⁰ with the chosen fields for this research. Table 2 presents the fields of interest included in the project data model and their description. I then further analysed the data model using a Pivot tool of Splunk, which enables quick design of reports with tables and data visualisations that present different aspects of a created data model. Pivot has a user interface, which I used to perform basic searches as an alternative to the use of SPL. (158) For more complex searches, I combined the Pivot tool with SPL. Moreover, for the composition of graphs with several time series, I created an event type⁴¹ with the categories of events of interest, which was then included in the data model.

Table 2. Fields of interest included in the project data model and their descriptions. (160)

Field name	Field description
interaction.author.id	The author's ID on the service from which they generated a post.
interaction.author.username	The username of the author of a post.
interaction.content	The main content of a post.
interaction.created_at	The time and date of a post creation.
interaction.geo.latitude	The geographical latitude of a geo-enabled post.
interaction.geo.longitude	The geographical longitude of a geo-enabled post.
interaction.hashtags	A list of the hashtags found in a post.
interaction.id	The identification number of a post
interaction.link	A link to an individual post.
interaction.mentions	List of users/usernames mentioned in an interaction.
interaction.source	The application that was used to generate and send a post.
language.tag	Two-character language code that identifies the language in which a post is written.
links.title	The title of the page the link points to.
links.url	The full URL of the page (resolved to the final page).
twitter.id	Identification number of a tweet.
twitter.in_reply_to_screen_name	The Twitter username of the author a tweet replies to (if it is a reply).

⁴⁰ "A data model is a hierarchically structured search-time mapping of semantic knowledge about one or more datasets. It encodes the domain knowledge necessary to build a variety of specialized searches of those datasets. These specialized searches are used by Splunk Enterprise to generate reports for Pivot users". (157)

⁴¹ "A user-defined field that represents a category of events". (159)

twitter.retweet.count	The total number of retweets for this tweet.
twitter.retweet.id	Identification number of a retweeted tweet.
twitter.retweet.mentions	List of Twitter usernames mentioned in a tweet that was retweeted.
twitter.retweet.user.id	The ID of the retweeting user.
twitter.retweet.user.screen_name	The retweeting user's Twitter username.
twitter.retweeted.id	The unique ID of the tweet that was retweeted.
twitter.retweeted.user.followers_count	The number of followers the retweeted user has.
twitter.retweeted.user.id	The Twitter ID of a retweeted user.
twitter.retweeted.user.screen_name	The retweeted user's Twitter username.
twitter.user.description	The Twitter user's biographical description.
twitter.user.followers_count	The number of followers the user has.
twitter.user.friends_count	The number of people an author follows.
twitter.user.id	Unique identification number of the Twitter user.
twitter.user.screen_name	A user's Twitter username.

4.5. Data cleaning

Although DataSift attempts to ensure 100% data delivery, due to network latency, system latency or other unforeseen issues, duplicates can occur as the provider repeats delivery attempts (even if partially successful) in cases when the delivery was interrupted. For the same reason, gaps may also occur in delivery. The data provider places the responsibility on the receiving party to handle duplicates or identify gaps. Therefore, I checked for errors and consistency in the distribution of the events and fields in the dataset by examining the temporal frequency distributions, means, and medians. By doing so, I noted a high number of duplicate events spread across the dataset. Using Splunk I, with the assistance of Splunk team, was able to identify and eliminate the duplicates and fill recognised gaps, bringing the dataset down to 60,511,402 events. This dataset acted as the final dataset for my study, from which I built the final accelerated data model. However, several gaps still remained across the dataset that could not be refilled due to the absence of events in raw data. Nevertheless, this was an expected limitation while working with

large datasets of this kind, the identified gaps were short and, in terms of the time period they covered, they did not exceed the length of two hours.

4.6. Description of the full dataset

I described the full dataset by the number of days it covered, the number of tweets and retweets, language distribution of tweet texts, the number of unique user accounts, the number of unique URLs shared, the number of tweets and retweets containing at least one URL, and the number of users that shared at least one URL. I also calculated the average number of tweets per day. Further, I identified peaks (i.e. days with the highest number of tweets) and described them using the same metrics as for the full period.

I also examined the geography of tweets in the dataset. I identified and mapped the geo-tagged tweets using the Splunk world map visualisation tool in order to demonstrate the geographical distribution of tweets. Only original tweets in the dataset (not retweets) could potentially be geo-tagged.

4.7. Identification of study sub-periods

For the purposes of this research, I focused on the temporal distribution of tweet frequency by day within the dataset. Due to a long temporal period of the dataset coverage, I developed a way to break it down into smaller sub-periods for the further analysis. I used the average temporal tweet frequency distribution as a base for this division. Thus, based on the striking difference in the tweet frequency in particular periods, I broke the dataset into the five following sub-periods (Figure 15):

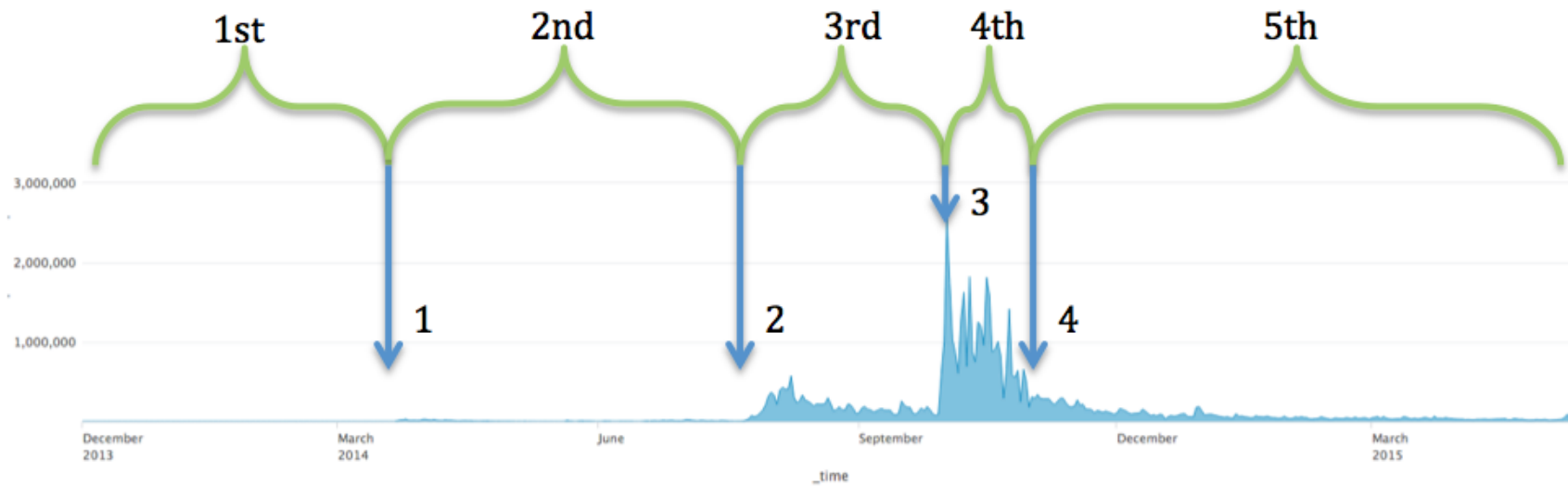
1. 1 December 2013- 21 March 2014
2. 22 March 2014 – 22 July 2014
3. 23 July 2014 – 29 September 2014
4. 30 September 2014 – 31 October 2014
5. 1 November 2014 – 12 May 2015

I chose the borderline dates for sub-periods due to the striking difference between the tweet frequency in those two days as demonstrated in Figure 15. This was observable on the overall temporal tweet frequency distribution graph. The overall level of tweets per day was visually at the same level within a sub-period with the exception for the 4th sub-period. This sub-period was characterised by dramatic changes in tweet frequency throughout October 2014 with significant increases and drops changing each other. However, on 31 October 2014 the last dramatic drop happened (178,757 tweets were posted, which was followed by a significant increase in tweet frequency on 1 November 2014 (313,406 tweets were posted)), which, in turn, switched to a gradual decrease in tweet frequency throughout the 5th sub-period.

Although the first reported case of EVD dates back to 26 December 2013, the temporal tweet frequency distribution from 1 December 2013 to 26 December 2013 (before the beginning of the outbreak) was not visually different from that in the period from 26 December 2013 to 21 March 2014. Therefore, I assumed that the Twitter Ebola conversation was initiated on 22 March 2014, and treated the 1st study sub-period as a baseline.

In order to make an attempt to identify the reason for the difference in the temporal tweet frequency distribution, I compared this with the WHO data on EVD reported cases and deaths, as well as with the number of new EVD cases per week in the most affected countries. Moreover, in order to check whether Twitter Ebola conversation may correlate with Ebola related news events, I matched the borderline dates (22 March 2014, 23 July 2014, 30 September 2014, 1 November 2014) between the sub-periods with the Ebola

Figure 15. The overall temporal tweet frequency distribution (by day) and the study sub-periods identified in the dataset.



Sub-period	Dates	Number of tweets on a borderline date/date	
1	1 December 2013- 21 March 2014		479/ 21 March
2	22 March 2014 – 22 July 2014	7,728/ 22 March	3,603/ 22 July
3	23 July 2014 – 29 September 2014	17,950/ 23 July	94,137/ 29 September
4	30 September 2014 – 31 October 2014	581,817/ 30 September	178,757/ 31 October
5	1 November 2014 – 12 May 2015	313,406/ 1 November	

news events that happened on/around these dates. I identified Ebola news events via HealthMap⁴², an online tool that gathers data from different sources in order to demonstrate the current global state of infectious diseases.

4.8. Description of study sub-periods

I described each of the 5 study sub-periods by length (number of days) and by the same metrics as the full dataset period. In addition, I identified and further explored the peak days (the days with the highest number of tweets within each sub-period) in the temporal tweet frequency distribution within a sub-period.

In order to make an attempt to quickly and efficiently identify the top topics of discussion within the peak days, I decided to explore the titles of the articles shared on this day as an alternative to a traditional content analysis of tweets. The latter would require significant time and human resources to individually read and code a sample of tweets for content. Looking at the most frequently shared links meant that it was possible to know that the same important information was being shared without having to code each share. Thus, I used the links augmentation in order to explore the titles of the top 20 shared URLs among the tweets within each peak. I translated the URL titles that were not in English on Google Translator⁴³. I then counted the number of URLs that contained identifiable article titles⁴⁴ and calculated the number of URLs among the top 20 shared URLs having the article titles on the same topic.

⁴² <http://www.healthmap.org/ebola/#timeline>

⁴³ <https://translate.google.co.uk/>

⁴⁴ By identifiable article links I mean the URL titles which included not only the title of the website they were posted on, but the title of news article.

4.9. Retweets

Retweets are reposts of someone else's tweets. Twitter provides a retweet feature, which enables its users to share a tweet of interest with all their followers. This feature brings additional information to tweet metadata. Thus, it includes the information about the original tweet that was retweeted and its sender as well as the information about the retweet and its sender. Before the retweet feature was officially introduced, Twitter users used to type "RT" at the beginning of a tweet to indicate that this is a retweet. Some of the current Twitter users still prefer to use this option. Moreover, some of the retweets are posted through third-party applications that have their own built-in version of retweeting. These retweets are not treated and identified as official retweets on Twitter and are missing additional information that Twitter includes on retweets. (161) Therefore, I did not treat these tweets as retweets in the study dataset.

I explored the retweets that were posted with the use of a specific Twitter retweet feature in order to find out what were the most retweeted accounts and the most retweeted tweets within the study dataset. I identified the number of retweets, the temporal frequency distribution of retweets, the number of unique retweeting and retweeted accounts. I also identified the top 20 retweeted accounts, and the top 10 retweeted tweets. For that, I identified the highest number of retweets achieved by these tweets over the whole study period. I then manually categorised the top 20 retweeted accounts and the accounts that posted the top 10 retweeted tweets into the following 10 account types that were identified using an inductive approach:

1. Celebrity
2. (Inter)governmental organisation/NGO⁴⁵
3. News
4. Other verified

⁴⁵ Non-governmental organisation

5. Humour
6. Infectious disease tracking
7. Automated tweet aggregator
8. Twitter celebrity
9. Individual
10. Other non-verified

The definitions and the algorithm that I developed and used for the categorisation of the Twitter accounts are presented in Table 3 and Figure 16.

Table 3. Definitions of account types for Twitter account categorisation with examples.

Type of account	Definition	Example
Celebrity	Verified official Twitter account of a celebrity or a public figure known outside of the Twittersphere.	@onedirection 1D's new track Drag Me Down - out now http://smarturl.it/1DdmdIT http://smarturl.it/1DdmdSP
(Inter)governmental organisation/NGO	Verified official Twitter account of a health related organisation (governmental, intergovernmental, non-governmental).	@WHO Official Twitter account of the World Health Organization, the United Nations' health agency
News	Verified official Twitter account of a news organisation.	@nytimes Where the conversation begins. Follow for breaking news, special reports, RTs of our journalists and more from http://NYTimes.com
Other verified	Verified Twitter account not categorised as "Celebrity", "(Inter)governmental organisation/NGO", or "News organisation".	@midnight Late night game show thing hosted by @nerdist on @ComedyCentral! http://instagram.com/atmidnightcc
Humour	Non-verified Twitter account that posts humour content (according to the user description).	@WorldStarFunny Posting the funniest content across the web. *Original Fan/Parody Account* NOT Affiliated With @WORLDSTAR or vine! worldstarf@gmail.com Warning:18+ content
Infectious disease tracking	Non-verified Twitter account that specialises in tracking infectious diseases (according to the user description).	@GertvanderHoek Tracking infectious diseases & other health threats
Automated tweet aggregator	Non-verified Twitter account not categorised as "Humour" or "Infectious disease tracking" that has posted more than 1,000 "Ebola" tweets over a study period.	@233liveOnline Delivering the latest news from Ghana and the world.
Twitter celebrity	Non-verified Twitter account not categorised as "Humour", "Infectious disease tracking" or "Automated tweet aggregator" and having at least 1,000 followers across the whole study period.	@medschooladvice Duke pediatric sleep doctor. Author of EVERYTHING I LEARNED IN MEDICAL SCHOOL. Advice giver-outer.

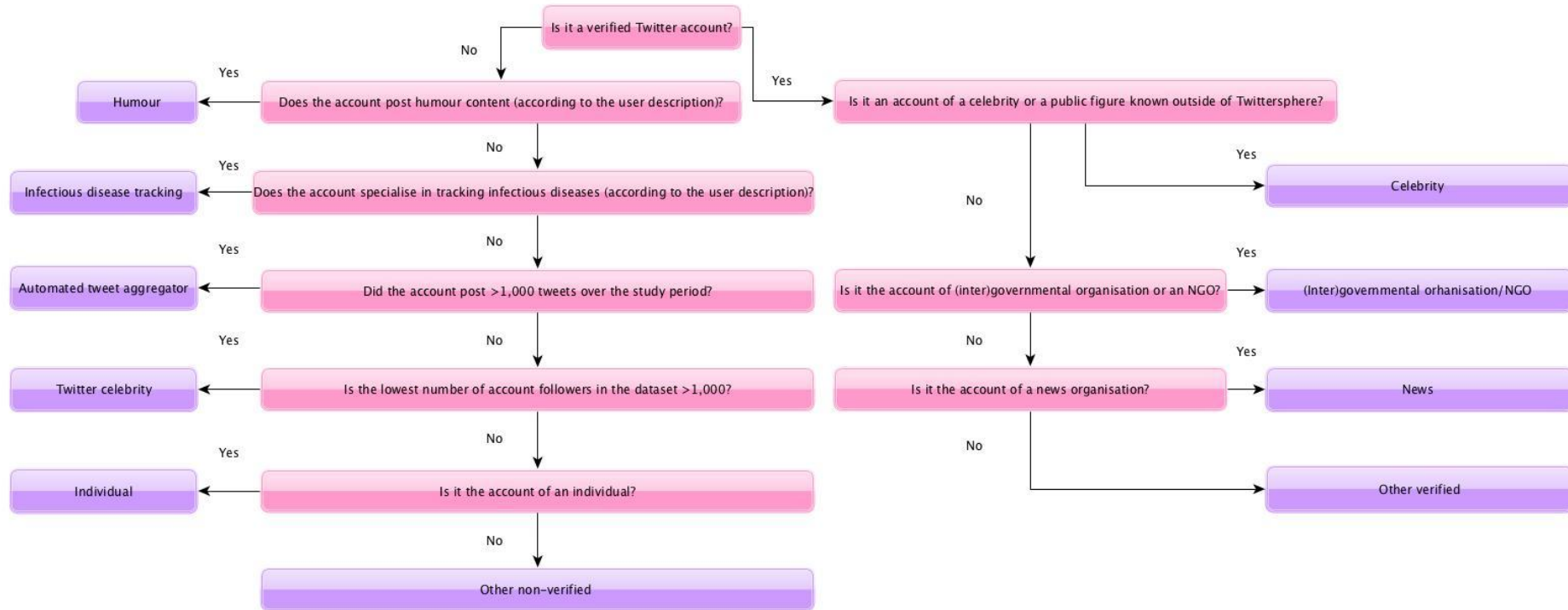
Individual	Non-verified Twitter account of an individual having less than 1,000 followers at least once during the study period.	@blossom_of_snow I have confidence in sunshine.
Other non-verified	Non-verified Twitter account not categorised as “Humour”, “Infectious disease tracking”, “Automated tweet aggregator”, “Twitter celebrity”, or “Individual”.	@olofofounded We create Awareness and Promotion for Your Brand, contact us now:- PIN 2B5506AB or DM

I further calculated the proportion of retweets received by each type of account out of the total number of retweets received by the top 20 most retweeted accounts.

Owing to privacy issues, I decided not to publish tweets that have been deleted by the time this dissertation is submitted, or those posted by users that have fewer than 1,000 followers (as I did not consider them as public figures). These tweets and Twitter users were still a part of the analysis, but I replaced their details by “****” symbol.

I applied the same methodology to each of the 5 sub-periods separately. I described the summary of the account categorisation by the most retweeted account types overall, within each sub-period, and in comparison to the baseline sub-period.

Figure 16. The algorithm for Twitter account categorisation in the dataset.



4.10. Mentions and replies

“A mention is a tweet that contains another user’s @username anywhere in the body of the tweet”. (138) Mentions are seen by mentioned users and by the followers of a sender. They can be searched for in a search box using the term “@username”.

Twitter users may start or join a conversation with other Twitter users by posting them a reply or by mentioning them. A reply is a response to another Twitter user by starting a message with “@username” of a person the reply is sent to. Twitter provides an official reply button for that. (138) The replies are seen in the timelines of the sender, the recipient, and the followers of both of them. They are also seen on the page of the sender in the “tweets & replies” section. Although a @username in a reply is also considered as a mention, the replies posted via an official reply button contain additional information that identifies them as a reply.

Mentions in the dataset also contained retweets as each official Twitter retweet in the dataset starts with “RT @username”. Therefore, the tweets including “any mentions” in the dataset contained replies, mentions, and retweets. Using SPL it was possible to retrieve mentions without retweets among them, so that these tweets included only mentions and replies. It was also possible to study replies independently.

I mapped the temporal tweet frequency distribution containing tweets, “any mentions”, replies, and mentions without retweets, and counted their number in the whole dataset. Mentions without retweets and replies were of a further interest for this study, as they demonstrate the direct engagement between Twitter users. I recorded the number of “any mentions”, “any mentioned” and “any mentioning” accounts. I further calculated the number of mentions, unique mentioned, and mentioning accounts and categorised the top 20 mentioned accounts into 10 account types using the same definitions and the algorithm presented in Table 3 and Figure 16. Although some of the accounts were mentioned and

received replies, they did not have any tweets posted within the study dataset. Therefore, it was impossible to identify some of the features needed for the categorisation (i.e. user description, the number of followers they had, and whether they were verified by Twitter), and they were marked as “n/a”. I applied the same methodology for each of the 5 sub-periods analysis. Furthermore, I performed the same analysis for the investigation of replies. Finally, I described the summary of the account categorisation by the most mentioned account types overall, within each sub-period, and in comparison to the baseline sub-period.

4.11. Retweets, mentions, and replies of (inter)governmental organisation/NGO accounts

I analysed the categorisation results of the top 20 retweeted accounts, the top 20 mentioned accounts, the top 20 accounts that received the highest number of replies, and the accounts that posted the top 10 retweeted tweets by looking at the presence of (inter)governmental organisation/NGO accounts among them. I identified the organisations owning this type of accounts within the study period and each of the 5 sub-periods and described their presence.

4.12. WHO and CDC Twitter Ebola activity

In an effort to explore the ways in which the WHO and CDC used Twitter during the West African Ebola outbreak and the level of engagement they achieved with the general Twitter population, I examined the accounts of the WHO (@WHO) and CDC (@CDCgov).

4.12.1. WHO and CDC Ebola tweets and their content

I identified the number of tweets posted by the WHO and CDC over the full study period, and examined the temporal tweet frequency distribution. I then conducted the

content analysis of tweets using an inductive approach, and grouped the tweets into the following categories:

1. Statements, announcements, reports
2. Press release, press conference, press interviews
3. Retweets
4. Tweeting from official events
5. Personal story, feature story
6. Responses and Twitter chats
7. Public health messages for the general public
8. Public health messages and information for health workers

The definitions of each category that I developed and used are presented in Table 4.

Table 4. Categories of the WHO and CDC messages, definitions, and examples.

Tweet category	Definition	Example
Statements, announcements, reports	The tweets that included a text or an URL referring to official organisational statements, announcements, reports, or scientific articles.	@WHO. WHO declares the end of the #Ebola outbreak in #Liberia Liberia #StaysVigilant
Press/news release, press conference, interviews	The tweets that included a text or an URL referring to official press/news releases; the tweets that were posted featuring a press conference or an interview of an official organisation representative; tweets that included other materials related to these events.	@WHO. Health partners unite to build stronger #healthsystems in #Ebola-affected countries http://t.co/cNpz4aJ74w
Retweets	The retweets or modified tweets, originals of which were posted by Twitter accounts, other than any other official accounts of the organisation.	@CDCgov.RT @WhiteHouse: "The United States will continue to help lead the global response in West Africa." —President Obama on #Ebola: http://t.co/sbnrch3Zxg
Tweeting from official events	Tweets that were posted featuring official events of the organisation or where an official organisational representative participated / tweets that included other materials related to these events.	@WHO. WHO Executive Board Special Session on #Ebola starts now. Follow live webcast here http://t.co/3u0sVeDTbq #EB136

Personal story, feature story	Tweets that included a text or an URL referring to personal or feature stories posted on the organisation website/other tweets related to personal or feature stories.	@CDCgov. CDC #DiseaseDetective Rupa trained Guinean medical workers who screen river passengers for symptoms of #Ebola. http://t.co/cHzyiHIj9q
Responses and Twitter chats	Tweets that were posted as direct responses to Twitter users/tweets of the Twitter chats/other tweets related to Twitter chats	CDCgov.@akinmaven Ebola is not spread through the air. Mosquitos or other insects are not thought to be able to transmit Ebola. #EbolaChat
Public health messages for the general public	Tweets, containing public health messages and targeting the general public	@WHO.Semen may be infectious with #Ebola virus for as long as 3 months after onset of symptoms http://t.co/TfWE5BIfTW http://t.co/N3IWE5yqmo
Public health messages and information for health workers	Tweets, containing public health messages and targeting health workers	CDC wants healthcare workers to be safe & prepared. Learn CDC recommended PPE for #Ebola: http://t.co/qXbUhzUpbB http://t.co/bic6MwWqsw

I examined the prevalence of different categories of tweets throughout the study period and 5 sub-periods and identified striking peaks in the temporal tweet frequency distribution (40 or more tweets per day for the @WHO account and 30 and more tweets for the @CDCgov account). I recorded the category distribution of tweets on these days and manually identified the main topic of tweets.

4.12.2. WHO and CDC retweets

I identified the number of retweets of the WHO and CDC tweets and the number of unique Twitter retweeting users over the study period, as well as the retweet temporal frequency distribution (by day) and the peak days in it.

Furthermore, I identified and categorised the top 10 most retweeted tweets of these accounts. I subsequently recorded the highest number of retweets for these tweets in the full dataset, as some of the tweets could have caught Twitter population attention several days after the original tweet creation date. The focus of this analysis was to identify the

most retweeted types of tweets. I therefore recorded their creation date (with the study sub-period), the number of retweets and the tweet type.

I analysed further the “Public health messages for the general public” tweet category. I manually extracted these tweets from the dataset, and recorded the number of their retweets. As more than 50% of retweets are usually posted within the first hour after a tweet creation, (140) to capture most of the “Public health messages for the general public” tweet retweets and to make the results comparable, I decided to look at the number of retweets that were posted within the first three days after the original tweet creation. It allowed me to track the temporal change of the level of public interest in public health information provided by these accounts throughout the study period. I further identified the most retweeted tweets (retweeted more than 1,000 times over the whole study period) for each account and recorded the number of their retweets.

4.12.3. WHO and CDC mentions and replies

I identified the number of the @WHO and @CDCgov mentions without retweets and the number of unique Twitter mentioning users over the full study period, as well as the temporal mention frequency distribution (by day) and the peak days in it. I performed the same for the replies to the @WHO and @CDCgov accounts. I recorded the number of geo-tagged mentions and demonstrated their geographical distribution using the Splunk World Map tool.

4.12.4. Followers of the WHO and CDC accounts

To identify the number of the account followers I used the data that came with each of the tweets posted by the WHO or CDC. Therefore, there were gaps in the data on the days when the WHO or CDC did not post any Ebola tweets. I identified the lowest and the highest number of the account followers in the dataset. I mapped the graphs presenting the

growth in the number of followers throughout the full study period. I visually identified the dates when the growth level was changing and recorded the events that happened on these dates.

4.13. Research ethics

The Research and Ethics Committee of the Oxford Internet Institute, University of Oxford, reviewed the research and approved it on 10 August 2015 with the reference number: OII C1A 15-045 (Appendix 5).

Chapter 5. Results

5.1. Description of the full dataset

The full dataset included 60,511,402 tweets, covering the period of 528 days from 1 December 2013, 8:00:00 UCT to 12 May 2015, 06:00:00 UCT. The average number of tweets per day was 114,605. 48.3% of the tweets in the dataset were retweets resulting in the total number of 29,227,230 retweets. The number of unique user accounts observed in the dataset was 12,185,699. 5,063,348 unique URLs were shared in 32,970,798 tweets (54.5% of the overall number of tweets) by 5,505,965 users (45.2% of the overall number of unique users).

The dataset included tweets in 149 languages with the top 10 languages accounting for 96.4% of the tweets (58,312,020). English was the most prevalent language among the top 10 languages (83.4%) and this was followed by Spanish (5.4%) and Portuguese (4.2%) (Table 5).

Table 5. Top 10 tweet text languages prevalent in the dataset.

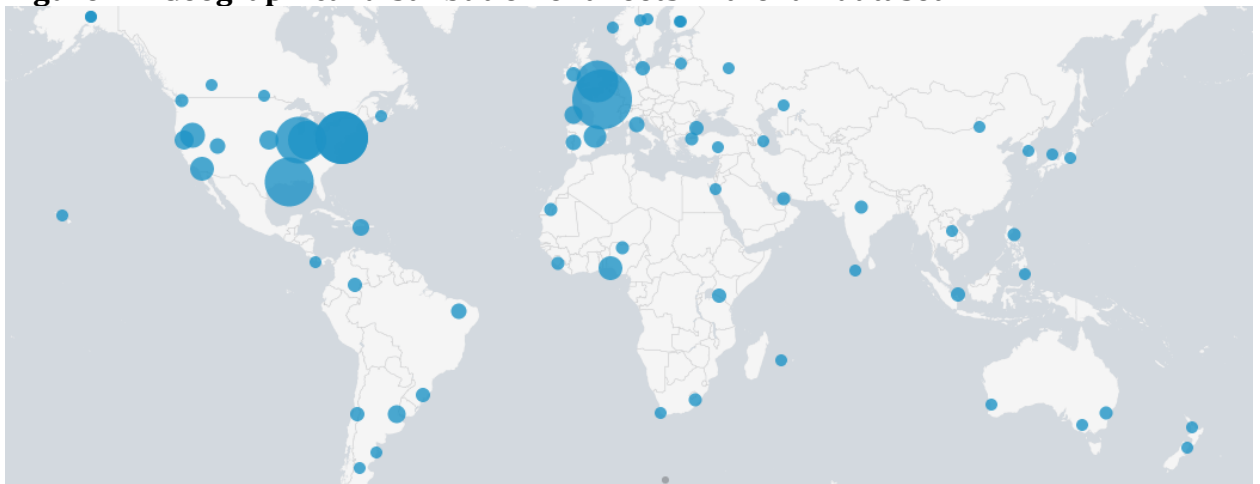
No	Language	Prevalence among the top 10 languages
1	English	83.4%
2	Spanish	5.4%
3	Portuguese	4.2%
4	French	3.0%
5	Indonesian	1.3%
6	Italian	0.8%
7	Turkish	0.6%
8	Dutch	0.6%
9	German	0.5%
10	Malay	0.3%

2 October 2014 was the peak day in the dataset with 2,658,873 tweets posted by 1,358,320 users, among these tweets 1,795,868 (69.9%) were retweets. On this day 80,466 unique URLs were shared in 550,736 tweets (21.4% of the number of the tweets posted on

that day) by 282,059 users (20.8% of the total number of the unique users who posted on that day).

1.2% of the overall number of tweets in the dataset (739,709) were geo-tagged. However, as only original tweets⁴⁶ in the dataset could potentially be geo-tagged, 2.4% of them actually contained geo-coordinates. Figure 17 demonstrates the geographical distribution of these tweets, with most coming from North America and Europe.

Figure 17. Geographical distribution of tweets in the full dataset.

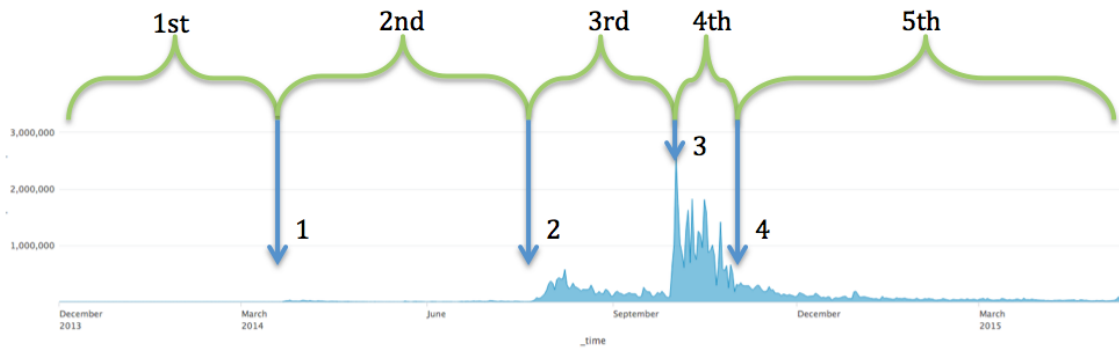


5.2. Study sub-periods and their correspondence with the Ebola infection rate and Ebola news events

I broke down the full study period into five shorter sub-periods based on the temporal tweet frequency distribution as demonstrated in Figure 18.

⁴⁶ Tweets without retweets

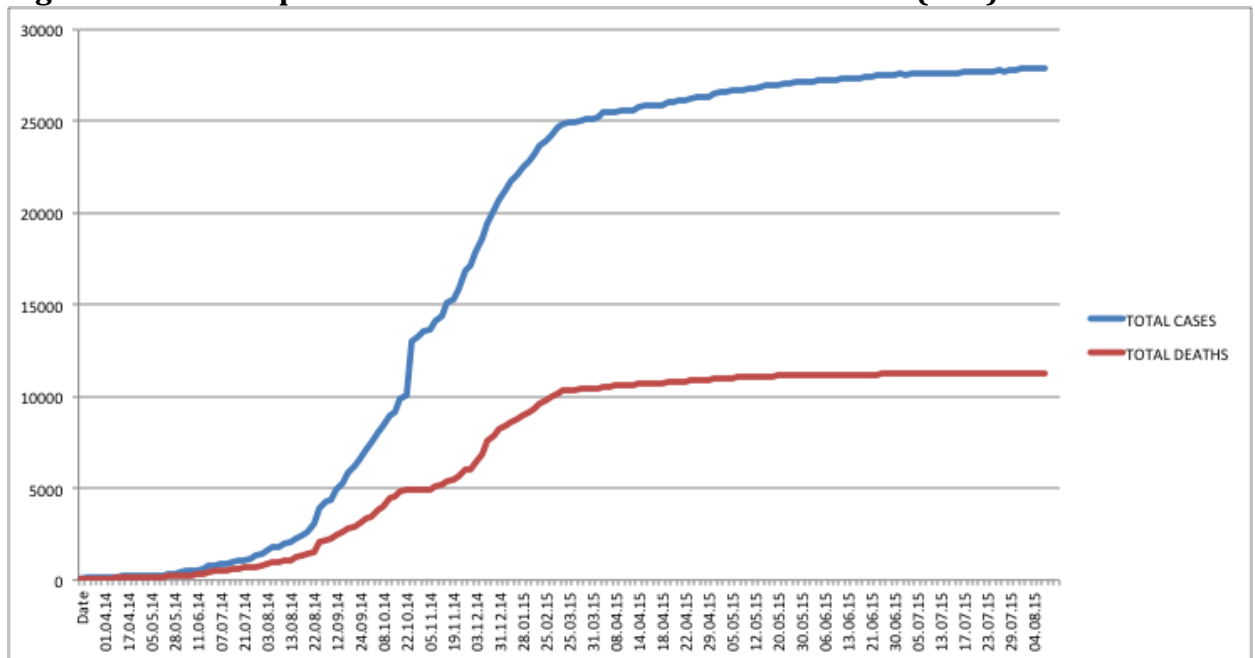
Figure 18. Overall tweet frequency distribution (by day) and the study sub-periods identified in the dataset.



Sub-period	Dates	Number of tweets on a borderline date/date	
1	1 December 2013- 21 March 2014	479/ 21 March	
2	22 March 2014 – 22 July 2014	7,728/ 22 March	3,603/ 22 July
3	23 July 2014 – 29 September 2014	17,950/ 23 July	94,137/ 29 September
4	30 September 2014 – 31 October 2014	581,817/ 30 September	178,757/ 31 October
5	1 November 2014 – 12 May 2015	313,406/ 1 November	

The temporal tweet frequency distribution did not seem to correspond with the total reported Ebola cases and deaths number over time (Figure 19).

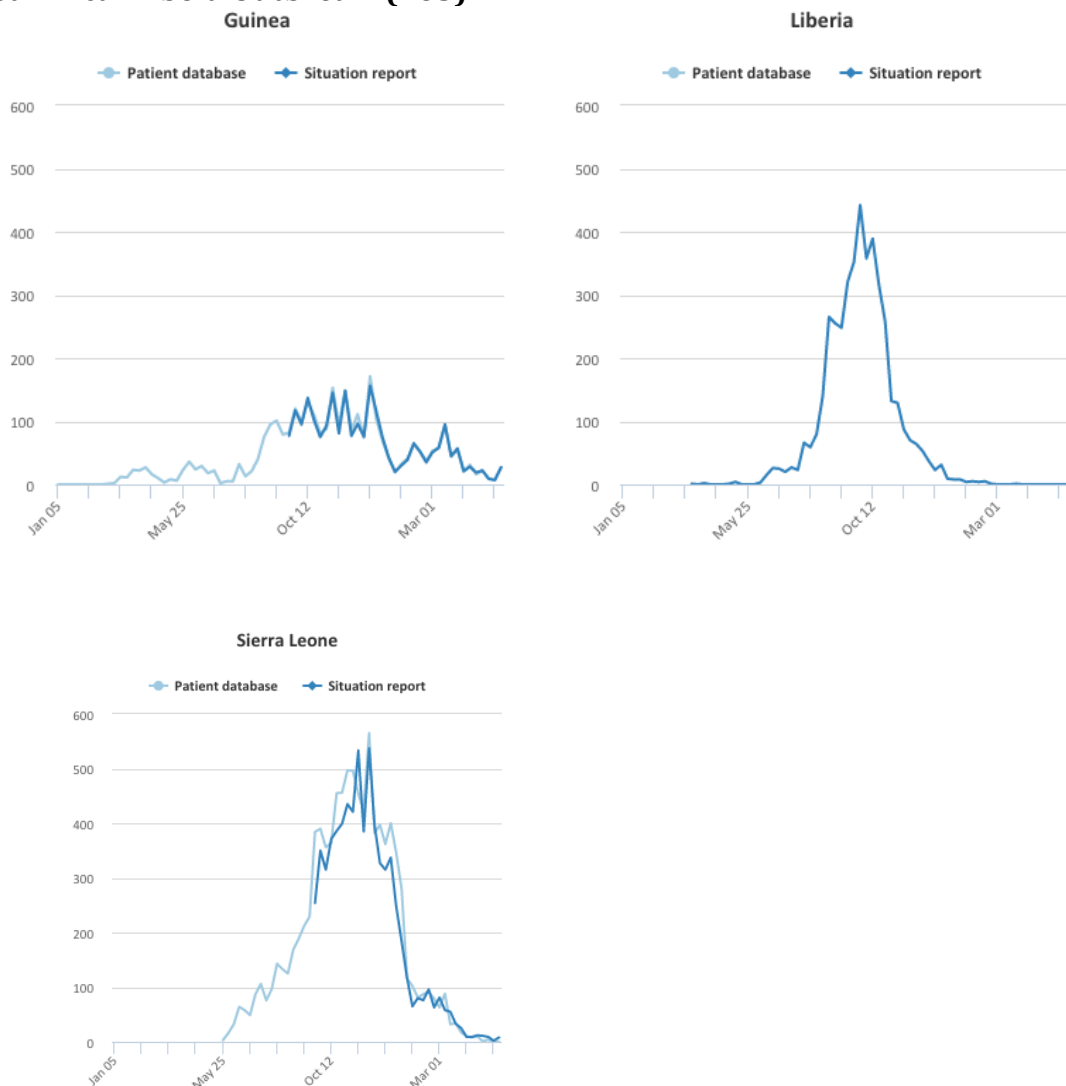
Figure 19. Total reported cases of deaths from EVD over time. (162)



Neither did it seem to correspond with the Ebola infection rate over time in the most affected countries (Guinea, Sierra Leone, or Liberia). However, it can be noticed that the

peak of infection rate in Guinea (September 2014 – February 2015), Sierra Leone (September 2014 – February 2015) and Liberia (September 2014 – November 2014) (Figure 20) resembled the shape of the temporal tweet frequency distribution in the dataset.

Figure 20. New confirmed Ebola cases in Guinea, Liberia, and Sierra Leone during the West African Ebola Outbreak. (163)



I identified the Ebola key news events that happened on the borderline dates of the temporal tweet frequency distribution. Thus, on 22 March 2014, Guinean health official received confirmation that the mysterious haemorrhagic fever was caused by Ebola virus and the outbreak was officially announced on the WHO website on 23 March 2014 with 49 cases and 29 deaths reported. (164) From the temporal tweet frequency distribution, I assumed that this news event initiated the growth of the public interest in the topic, and,

therefore, Twitter Ebola conversation, acting as the beginning of the 2nd sub-period. On 25 July 2014, Nigerian health officials announced that the first EVD case was diagnosed in Lagos. (165) Nevertheless, it has been demonstrated that the number of Ebola tweets started to rise in Nigeria several days before the official announcement. (166) This event seemed to cause the start of the new wave of the Twitter Ebola conversation and the beginning of the 3rd sub-period. It was further amplified by the news about the evacuation of the two Americans infected with EVD from Liberia to the USA at the beginning of August 2014 and by the WHO declaration of the Ebola outbreak a Public Health Emergency of International Concern on 8 August 2014. On the third borderline date, 30 September 2014, the CDC announced the first diagnosed EVD case in the USA, which caused a dramatic increase in the tweet frequency and served as the beginning of the 4th sub-period. During October 2014 I observed the striking fluctuation of tweet frequency distribution, which seemed to be mostly related to the four Ebola cases diagnosed in the USA. I was unable to discern the reason for the appearance of the borderline between the 4th and the 5th sub-periods and a decrease in the tweet frequency in that sub-period, but it was likely due to the fact that the Western media stopped covering the outbreak intensively after all the four diagnosed cases in the USA in October 2014. This may have caused a gradual decrease of the public fear and anxiety, and the increase of the public indifference to the topic.

5.3. Description of study sub-periods

The 5th sub-period in the dataset was the longest, consisting of 193 days (36.6% of the full period), whereas the 4th sub-period was the shortest, consisting of 32 days (6.1%). The 1st sub-period represented 21% of the full dataset length, the 2nd sub-period of 23.3%, and, lastly, the 3rd sub-period of 13.1% (Table 6).

Table 6. Description of the study sub-periods by length.

Sub-period	Dates	Number of days	% of the full period length
1	1 December 2013- 21 March 2014	111	21.0%
2	22 March 2014 – 22 July 2014	123	23.3%
3	23 July 2014 – 29 September 2014	69	13.1%
4	30 September 2014 – 31 October 2014	32	6.1%
5	1 November 2014 – 12 May 2015	193	36.6%

The full description of the 5 sub-periods with comparison to the full study period is presented in Table 7. The 1st baseline sub-period had the lowest average daily number of tweets at 323. Despite the fact that the 4th sub-period was the shortest in the dataset, 52.3% of the tweets were posted then with an average of 989,089 tweets per day, which is 3,062-fold higher than in the baseline sub-period, followed by the 3rd sub-period that contained 21.8% of overall tweets with 191,272 tweets as a daily average, and then by the 5th (114,605 tweets per day) and the 2nd (7,022 tweets per day) sub-periods.

During the 1st sub-period tweets were posted by 19,733 unique users, with the average daily number of unique users at 178. In subsequent sub-periods the absolute and the average daily number of Twitter users was higher, reaching the highest number of unique users in the 4th sub-period with 8,388,524 unique users joining the conversation on Twitter (with the daily average number at 262,141 unique users, 1,473-fold with the baseline period).

The prevalence of tweets containing URLs was higher in the 2nd (75.3%), 3rd (68.8%), and 5th (64%) sub-periods in comparison with the 1st sub-period (55.5%), and was lower in the 4th sub-period (43.5%). Nevertheless, the daily average number of unique URLs shared was higher than in the 1st sub-period (62 URLs) in all the subsequent sub-periods, reaching the highest number in the 4th sub-period at 62,111. Moreover, the prevalence of the unique users sharing the URLs (among the number of unique users per period) was higher in the 2nd (41.7%), 3rd (54%), 4th (39%), and 5th (47.3%) sub-periods in comparison with that in the baseline sub-period (34.5%).

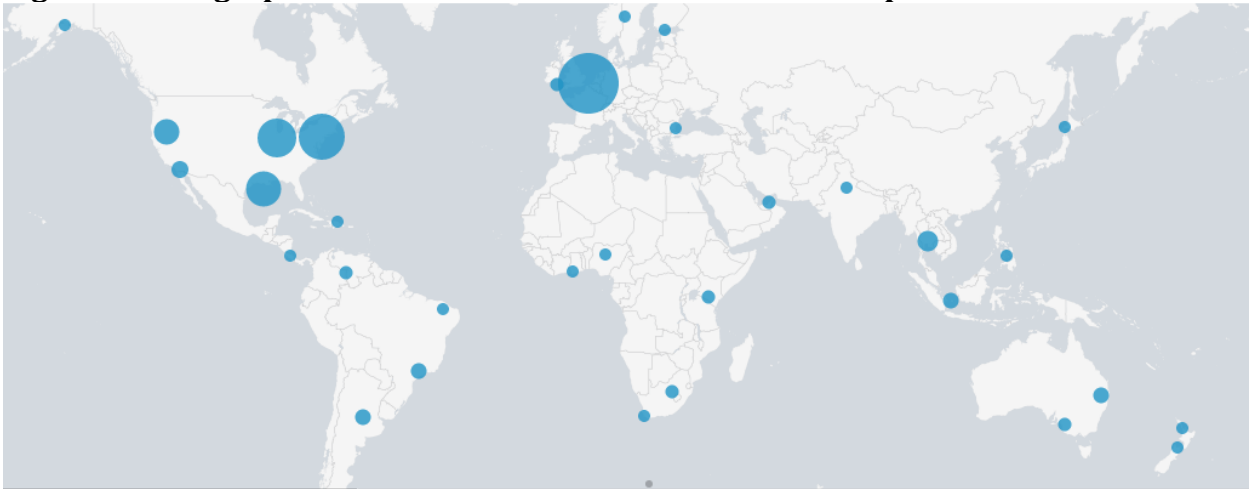
Table 7. Full description of the study sub-periods and the full period by the number of tweets, retweets, unique users and URLs.

Sub-period	Dates	Number of days (% of the number of days over the full period)	Number of tweets per sub-period (% of the number of tweets over the full period)	Number of retweets (% of the number of tweets per sub-period)	Average daily number of tweets per sub-period	Number of unique user accounts per sub-period (daily average)	Number of unique URLs shared per sub-period (daily average)	Number of tweets containing URLs over the sub-period (% of the number of tweets per sub-period)	Number of unique user accounts that shared at least one URL per sub-period (% of the number of unique user accounts per sub-period)
1	1 December 2013 - 21 March 2014	111 (21.0%)	35,823 (0.06%)	11,568 (32.3%)	323	19,733 (178)	6,877 (62)	19,877 (55.5%)	6,814 (34.5%)
2	22 March 2014 - 22 July 2014	123 (23.3%)	863,690 (1.4%)	303,647 (35.2%)	7,022	287,119 (2,334)	116,225 (945)	650,593 (75.3%)	199,795 (41.7%)
3	23 July 2014 - 29 September 2014	69 (13.1%)	13,197,756 (21.8%)	4,933,633 (37.4%)	191,272	2,713,528 (39,326)	1,467,123 (21,262)	9,084,373 (68.8%)	1,466,571 (54.0%)
4	30 September 2014 - 31 October 2014	32 (6.1%)	31,650,833 (52.3%)	17,661,398 (55.8%)	989,089	8,388,524 (262,141)	1,987,578 (62,111)	13,775,166 (43.5%)	3,269,823 (39.0%)
5	1 November 2014 - 12 May 2015	193 (36.6%)	14,763,300 (24.4%)	6,316,984 (42.8%)	76,494	4,731,119 (24,513)	1,576,218 (8,166)	9,440,789 (64.0%)	2,236,583 (47.3%)
Full period	1 December 2013 - 12 May 2015	528 (100%)	60,511,402 (100%)	29,227,230 (48.3%)	114,605	12,185,699 (23,079)	5,063,348 (9,590)	32,970,798 (54.5%)	5,505,965 (45.2%)

5.3.1. Geography of tweets

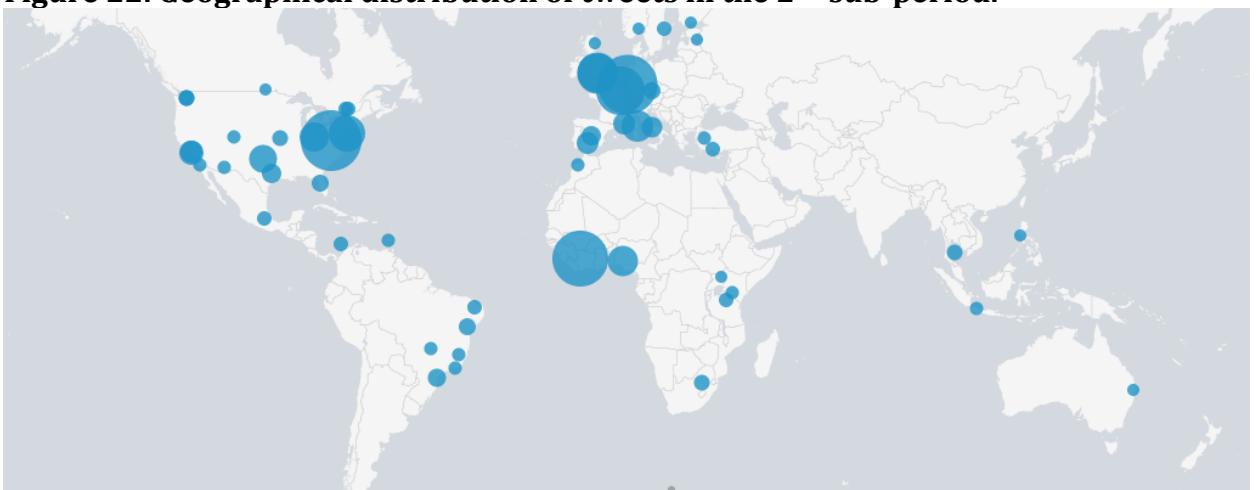
1,7% (602) of the overall number of tweets in the 1st sub-period were geo-tagged. Figure 21 demonstrates the geographical distribution of these tweets, with most coming from North America and Europe.

Figure 21. Geographical distribution of tweets in the 1st sub-period.



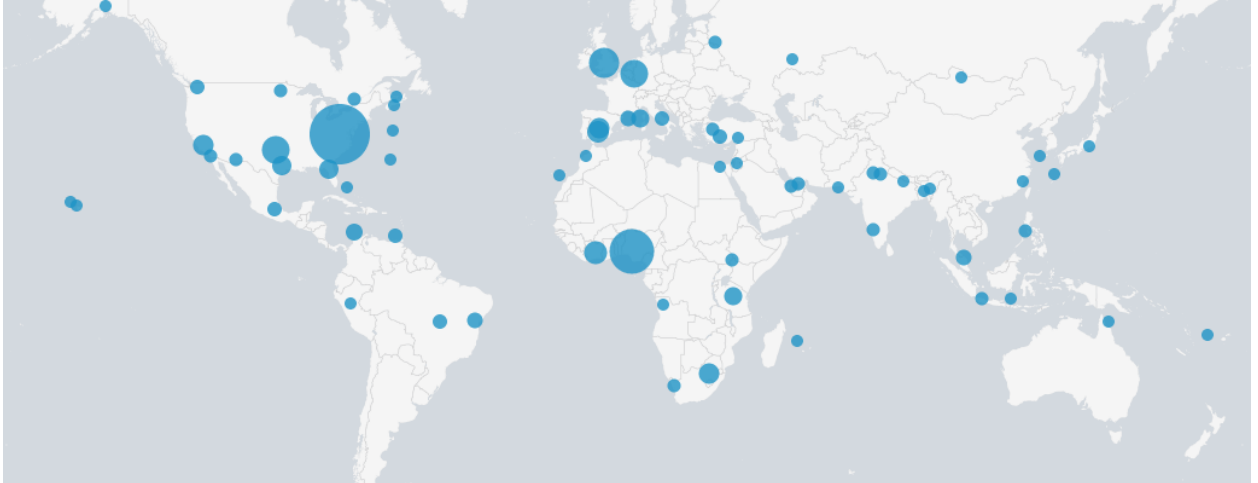
0.9% (7,436) of the overall number of tweets in the 2nd sub-period were geo-tagged. Figure 22 demonstrates the geographical distribution of these tweets, with most coming from North America, Europe (especially Western Europe), and West Africa (especially Guinea, Liberia, and Sierra Leone).

Figure 22. Geographical distribution of tweets in the 2nd sub-period.



1.1% (147,134) of the overall number of tweets in the 3rd sub-period were geo-tagged. Figure 23 demonstrates the geographical distribution of these tweets, with most coming from North America, Europe, and West Africa (especially Nigeria).

Figure 23. Geographical distribution of tweets in the 3rd sub-period.



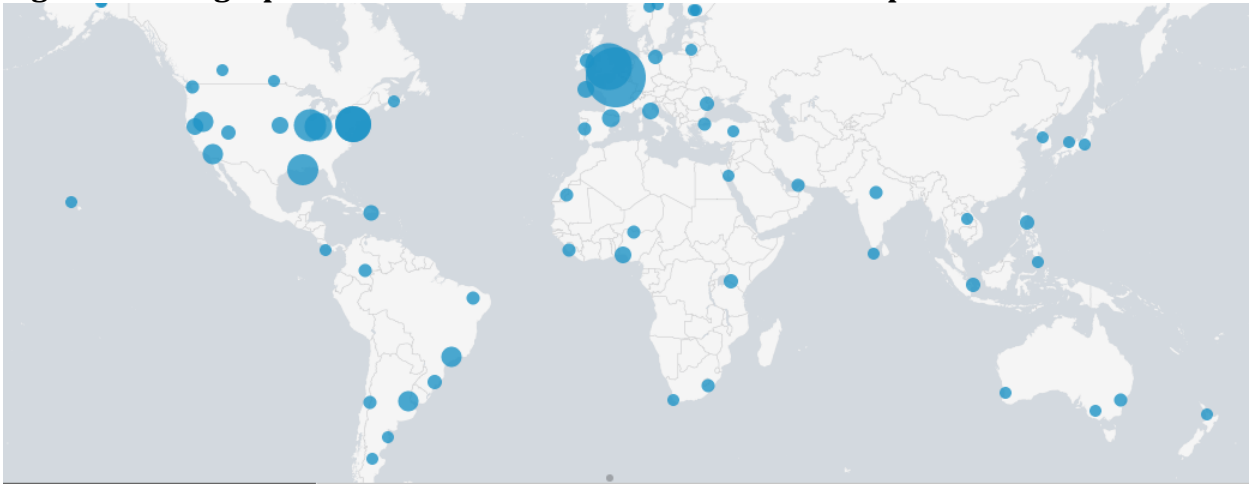
1.3% (418,242) of the overall number of tweets in the 4th sub-period were geo-tagged. Figure 24 demonstrates the geographical distribution of these tweets, with most coming from North America.

Figure 24. Geographical distribution of tweets in the 4th sub-period.



1.1% (166,295) of the overall number of tweets in the 5th sub-period were geo-tagged. Figure 25 demonstrates the geographical distribution of these tweets, with most coming from North America and Europe.

Figure 25. Geographical distribution of tweets in the 5th sub-period.



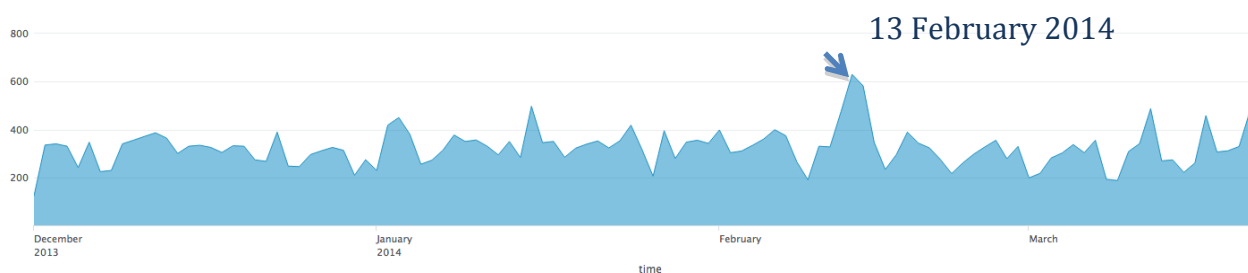
5.3.2. Peak days in study sub-periods and URLs on these days

I identified the following peak days for each of the sub-periods: 13 February 2014 (1st sub-period), 25 March 2014 (2nd sub-period), 8 August 2014 (3rd sub-period), 2 October 2014 (4th sub-period), 3 November 2014 (5th sub-period). Table 8 represents the comparative description of these peak days.

Table 8. Comparative description of peak days in each of the study sub-periods by the number of tweets, retweets, unique users, and unique links.

Sub-period	Dates	Average number of tweets per day in a sub-period	Peak day in a sub-period	Number of tweets in a peak day	Number of unique users in a peak day	Number of retweets in a peak day (% of the number of tweets in a peak day)	Number of unique URLs shared on a peak day	Number of tweets containing URLs on a peak day (% of the number of tweets in a peak day)	Number of unique users who shared at least one URL on a peak day (% of the number of unique users in a peak day)
1	1 December 2013 - 21 March 2014	323	13 February 2014	630	481	225 (46.8%)	137	498 (79.0%)	363 (75.5%)
2	22 March 2014 - 22 July 2014	7,022	25 March 2014	28,928	18,321	8,881 (30.7%)	4,280	23,220 (80.0%)	14,714 (80.3%)
3	23 July 2014 - 29 September 2014	191,272	8 August 2014	575,053	265,012	238,049 (41.4%)	54,169	339,735 (59.1%)	149,689 (56.5%)
4	30 September 2014 - 31 October 2014	989,089	2 October 2014	2,568,873	1,358,320	1,795,868 (69.9%)	80,466	550,736(21.4%)	282,059 (20.8%)
5	1 November 2014 - 12 May 2015	76,494	3 November 2014	341,143	233,091	164,528 (48.2%)	36,545	173,727 (50.9%)	96,050 (41.2%)

Figure 26. Temporal tweet frequency distribution (by day) during the 1st sub-period and the peak day.



13 February 2014 (Figure 26) was the peak day in the 1st sub-period, when 630 tweets (46.8% of them were retweets) were posted by 481 unique Twitter users. On this day 79% of the tweets contained 137 unique URLs shared by 75.5% of the unique tweeting users. Table 9 presents the titles of the top 20 unique URLs shared during the peak day in the 1st sub-period with the number of tweets containing them. As can be seen from Table 9, 18 of the 20 URL titles contained identifiable article titles; none of the URLs had a title suggesting the same topic.

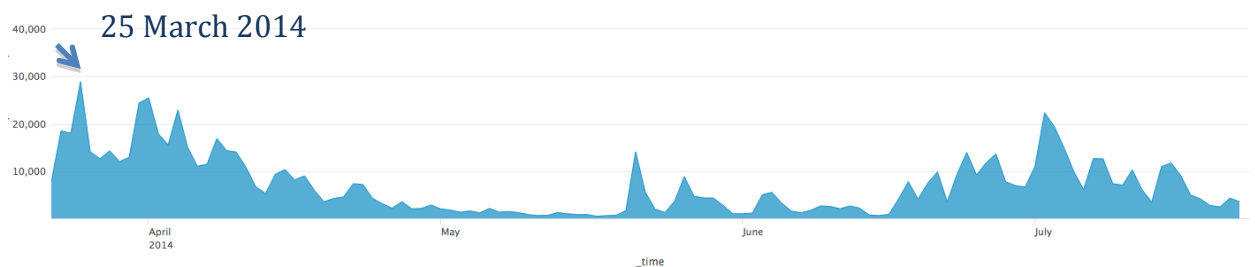
Table 9. Top 20 URLs shared on the 13 February 2014 (1st sub-period).

URL title	Number of times shared
U.S. launches new global initiative to prevent infectious disease threats - The Washington Post	201
Chimpanzee 'Mega-Culture' Discovered In Remote Congo Forest Might Be Africa's Largest	114
Euro Surveill. Possible pandemic threat from new reassortment of influenza A(H7N9) virus in China - FluTrackers	8
Jiangxi 1 new human infection with H10N8 avian flu - FluTrackers	7
KeywordResistanceFront.com Giving rogue low level government "intelligence" agencies something to read (snoop,hack,leak,scan,data mine,collect..)	7
H7N9 Cases February 13, 2014 (confirmed cases: 0, confirmed deaths: 0) (suspected cases: 0, suspected deaths:0) - FluTrackers	5
AdF.ly - shrink your URLs and get paid!	4
Chimpanzee 'Mega-Culture' Discovered In Remote Congo Forest Might Be Africa's Largest End The NRA!	4
Hyperbole and a Half: Boyfriend Doesn't Have Ebola. Probably.	4
Karma and Love BROWN CONDOR	4
null	4
H10N8 Case List 2013/2014 - FluTrackers	3
Highly pathogenic avian influenza H5N1, Cambodia (OIE, February 13 2014): 1 New Poultry Outbreak - FluTrackers	3

Highly pathogenic avian influenza H5N1, Vietnam (OIE, February 13 2014): 6 New Poultry Outbreaks - FluTrackers	3
Ireland Influenza Surveillance 2013-2014 - FluTrackers	3
Puerto Rico Dengue 2013 - FluTrackers	3
Spain Influenza Surveillance Reports 2014 - FluTrackers	3
WHO Risk Assessment of human infection with avian influenza A(H7N9) virus, as of January 13 2014 (edited) - FluTrackers	3
Chikungunya epidemie in het Caribisch gebied - FluTrackers (<i>Translation: Chikungunya epidemic in the Caribbean - FluTrackers</i>)	2
China - 2 more hospitals establish emergency procedures to accommodate respiratory patients on smoggy days - Zhangjiakou City, Hebei province - FluTrackers	2

Strikethrough indicates URL titles containing non-identifiable article titles.
 Italic indicates Google Translator translation to English.

Figure 27. Temporal tweet frequency distribution (by day) during the 2nd sub-period and the peak day.



25 March 2014 (Figure 27) was the peak day in the 2nd sub-period, when 28,928 tweets (30.7% of them were retweets) were posted by 18,321 unique Twitter users. On this day 80% of the tweets contained 4,280 unique URLs shared by 80.3% of the unique tweeting users. Table 10 presents the titles of the top 20 unique URLs shared during the peak day in the 2nd sub-period with the number of tweets containing them. As can be seen from Table 10, 17 of the 20 URL titles contained identifiable article titles, 10 of which were about a Canadian man who arrived from West Africa back to Canada and developed symptoms of a haemorrhagic fever resembling Ebola.

Table 10. Top 20 URLs shared on the 25 March 2014 (2nd sub-period).

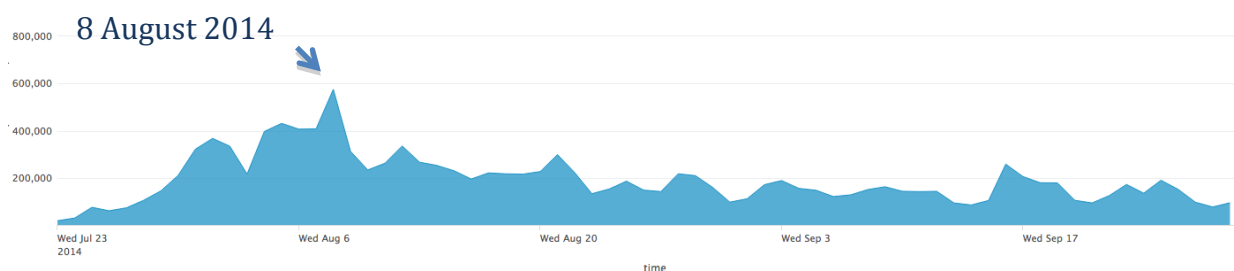
URL title	Number of times shared
null	2,828
BBC News - Canadian man in hospital with Ebola-like virus	1,536
BBC News - Guinea Ebola outbreak: Bats banned to curb virus	1,045
Canada probes possible Ebola case - CNN.com	905
Canada patient tests negative for Ebola - CNN.com	662
BBC News - West Africa on Ebola virus high alert	541
AdF.ly - shrink your URLs and get paid!	438
Suspected Ebola Case Is Detected In Canada	304
El Universal - El Mundo - Detectan posible caso de ébola en Canadá (Translation: El Universal - World - detect possible cases of Ebola in Canada)	261
From Africa to Canada: Deadly Ebola virus outbreak gets mobile — RT News	215
Teams Quarantine Victims in Guinea Ebola Outbreak - ABC News	209
Canada rules out suspected case of Ebola	201
Un cas suspect de fièvre Ebola détecté au Canada (Translation: A suspected case of Ebola detected in Canada)	184
Scientist who discovered Ebola frustrated by deadly Guinea outbreak Reuters	176
News from The Associated Press	160
West African nations scramble to contain Ebola threat Reuters	147
Man in Canadian Hospital With Ebola-Like Symptoms - ABC News	136
Sick man in Saskatoon tests negative for Ebola - Saskatchewan - CBC News	136
<i>59 muertes en Guinea por Ébola: Médicos Sin Fronteras SDP Noticias (Translation: 59 Ebola deaths in Guinea: MSF SDP News)</i>	129
Ebola mapping in Guinea: Humanitarian OpenStreetMap Team Mapbox	128

Strikethrough indicates URL titles containing non-identifiable article titles.

Italic indicates Google Translator translation to English.

Bold indicates the article titles of a similar topic.

Figure 28. Temporal tweet frequency distribution (by day) during the 3rd sub-period and the peak day.



8 August 2014 (Figure 28) was the peak day in the 3rd sub-period, when 575,053 tweets (41.4% of them were retweets) were posted by 265,012 unique Twitter users. On this day 59.1% of the tweets contained 54,169 unique URLs shared by 56.5% of the unique tweeting users. Table 11 presents the titles of the top 20 unique URLs shared on 8 August 2014 with the number of tweets containing them. As can be seen from Table 11, 14 of the 20 URL titles contained identifiable article titles, 8 of which had a title with the topic of the EVD being declared a Public Health Emergency of International Concern by the WHO.

Table 11. Top 20 URLs shared on the 8 August 2014 (3rd sub-period).

URL title	Number of times shared
null	22,290
The New York Times – Breaking News, World News & Multimedia	13,111
Update Your Browser Facebook	6,280
WHO: Ebola Outbreak Is International Public Health Emergency - NBC News	5,514
BBC News - Ebola 'an international emergency'	4,407
Ebola Facts	4,168
【閲覧注意】WHOがエボラ出血熱『緊急事態宣言』→感染するところなる (動画あり) コレスゴ! (Translation: [View Note: WHO is (with videos) of this section is made and Ebola haemorrhagic fever "Emergency Declaration" → infected / Koresugo!)	3,807
BBC News - WHO: Ebola 'an international emergency'	2,669
W.H.O. Declares Ebola in West Africa a Health Emergency - NYTimes.com	2,213
Ebola: U.S. orders diplomats' families to leave Liberia - CNN.com	2,141
WHO Frequently asked questions on Ebola virus disease	1,957
WHO declares Ebola epidemic an international health emergency Reuters	1,774
Instagram	1,538
AdF.ly – shrink your URLs and get paid!	1,336
'Don't Touch the Walls': Ebola Fears Infect an African Hospital - NYTimes.com	1,259
WHO WHO Statement on the Meeting of the International Health Regulations Emergency Committee Regarding the 2014 Ebola Outbreak in West Africa	1,169
Fujifilm Drug Eyed As Possible Treatment for Ebola Virus - Japan Real Time - WSJ	1,140
Ebola: uma emergência mundial – Infográficos (Translation: Ebola: a global emergency – Infographics)	1,120

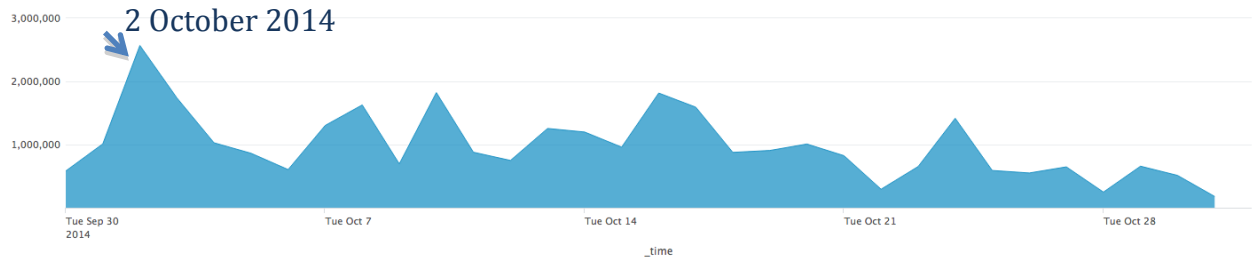
WHO: Ebola Outbreak is a Public Health Emergency	1,070
American Ebola Patient Kent Brantly: I Am Growing Stronger Every Day	1,056

Strikethrough indicates URL titles containing non-identifiable article titles.

Italic indicates Google Translator translation to English.

Bold indicates the article titles of a similar topic.

Figure 29. Temporal tweet frequency distribution (by day) during the 4th sub-period and the peak day.



2 October 2014 (Figure 29) was the peak day in the 4th sub-period, when 2,568,873 tweets (69.9% of them were retweets) were posted by 1,358,320 unique Twitter users. On this day 21.4% of the tweets contained 80,466 unique URLs shared by 20.8% of the unique tweeting users. Table 12 presents the titles of the top 20 unique URLs shared on 2 October 2014 with the number of tweets containing them. As can be seen from Table 12, 10 of the 20 URL titles contained identifiable article titles; however, none of them had a title suggesting the same topic.

Table 12. Top 20 URLs shared on the 2 October 2014 (4th sub-period).

Link title	Number of times shared
Ohhh shittt [2][2]	34,311
The New York Times – Breaking News, World News & Multimedia	26,568
-null	25,035
booty me down [2]	9,227
U.S. Ebola case; The fear, frustration and search grow - CNN.com	5,515
Lance Moore with touchdown dance [2] #fawkdawp YUNG POPPY	4,861
Biglexi	4,760
CNN Mobile	4,591
Bill Gates Tops Forbes 400 List for 21st Straight Year Video - ABC News	4,517
Ebola Overwhelming West Africa Communities - NYTimes.com	3,408
This the only nigga that can beat Floyd Mayweather	3,191

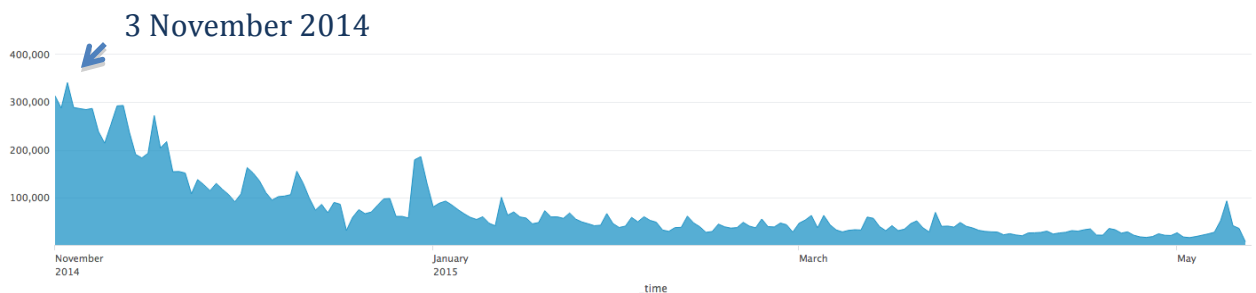
What are the countries most at risk Ebola spread? / Quali sono i Paesi più a rischio diffusione ebola? Eraliquida.com	3,050
BBC News - Ebola outbreak: 'Five infected every hour' in Sierra Leone	3,020
Thousands of children orphaned by Ebola abandoned - CNN.com	2,598
CDC Confirms First Case of Ebola Diagnosed in the U.S. The White House	2,567
REAL NIGGA ZAY on Twitter: "When you find out that ya ex died from Ebola http://t.co/aZL4W239ED"	2,559
I love my Ebola ❤️👏	2,383
👀👀👀👀 #wtf	2,300
AdF.ly - shrink your URLs and get paid!	2,293
BBC News - '80 had contact' with US Ebola man	2,246

Strikethrough indicates URL titles containing non-identifiable article titles.

Italic indicates Google Translator translation to English.

Bold indicates the article titles of a similar topic.

Figure 30. Temporal tweet frequency distribution (by day) during the 5th sub-period and the peak day.



3 November 2014 (Figure 30) was the peak day in the 5th sub-period, when 341,143 tweets (48.2% of them were retweets) were posted by 233,091 unique Twitter users. On this day 50.9% of the tweets contained 36,545 unique URLs shared by 41.2% of the unique tweeting users. Table 13 presents the titles of the top 20 unique URLs shared on 3 November 2014 with the number of tweets containing them. As can be seen from Table 13, 16 of the 20 URL titles contained identifiable article titles; however, none of them had a title suggesting the same topic.

Table 13. Top 20 URLs shared on the 3 November 2014 (5th sub-period).

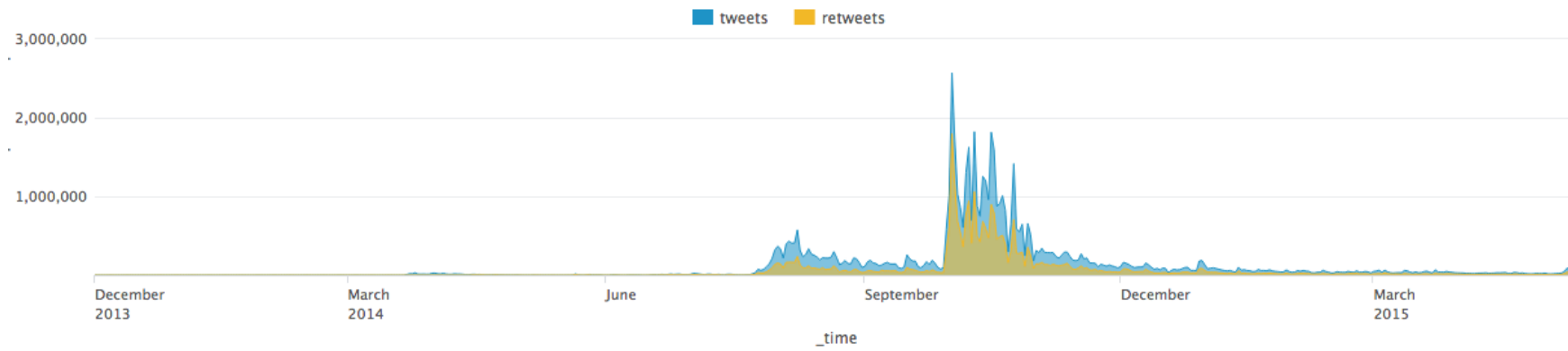
URL title	Number of times shared
null	7,210
UMA MENTIRA CHAMADA „EBOLA,, VEJAM !!! Veja na WEB (<i>Translation: A LIE CALLED,,, EBOLA SEE !!! See on the Web</i>)	4,066
Follow Ebola Updated every second & see what the #CDC & #WHO is not telling you about #Ebola	3,277
Shannon Ragland on Twitter: "Is St. Margaret Mary the dumbest school in Louisville? Teacher resigns because ebola scare. She was in KENYA! #ebola http://t.co/2pCakT0oZM"	2,262
Map: The Africa without Ebola - The Washington Post	1,937
When you lie and tell a bad ass kid they got Ebola👍👍👍	1,926
We Can Eradicate Malaria—Within a Generation Bill Gates	1,771
Can Obama Reboot? - Glenn Thrush and Carrie Budoff Brown - POLITICO Magazine	1,407
Ebola Cure Delayed by Drug Industry's Drive for Profit, W.H.O. Leader Says - NYTimes.com	1,303
UNICEF - Ebola	1,179
When you lie and tell a bad ass kid he got Ebola 🍑🍑🍑	1,166
<i>detikTV : Pilihan Cerdas Pencari Informasi (Translation: detikTV: Intelligent Choice Information Seekers)</i>	1,111
When the homie has EBOLA but u tryna cheer him up... (IB: Cody Ko)	1,102
El Universal - El Mundo - Enfermera logra acuerdo con Maine por cuarentena de ébola (<i>Translation: El Universal - World - Nurse Reaches Agreement with Ebola quarantine Maine</i>)	950
BioShield: Obama Diverted Its Funds From Fighting Ebola	827
No Such Thing as Ebola? Mobile Likes	816
Everyone in Ohio right now. #Ebola	810
Instagram	747
When you lie and tell a bad ass kid they got Ebola🍑🍑🍑 #Ebola #funny	705
Man jerking off to woman in public #Jamaica #Jamaican #YaadVineENT	663

Strikethrough indicates URL titles containing non-identifiable article titles.

Italic indicates Google Translator translation to English.

Bold indicates the article titles of a similar topic.

Figure 31. Tweet and retweet temporal frequency distribution (by day) in the dataset.



5.4. Retweets

The total number of retweets in the dataset was 29,227,230, which is 48.3% of the overall number of tweets. They were posted by 8,813,863 unique users, which means that 72.3% of the unique users in the dataset were retweeting other users. Moreover, 1,565,292 unique users were retweeted (12.8% of the overall number of unique users in the dataset). 42.5% of retweets contained URLs. Figure 31 demonstrates the temporal retweet frequency distribution (in yellow) compared to tweet frequency distribution (in blue) in the dataset, demonstrating that retweets followed the same pattern with tweets, presenting different proportion of them depending on the time period.

I categorised the top 20 retweeted accounts in the whole dataset into the following types:

- News - 9
- Humour – 6
- Inter(governmental) organisation/ NGO – 2
- Twitter celebrity – 2
- Automated tweet aggregator – 1

Figure 32 demonstrates the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the dataset.

Figure 32. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the dataset.

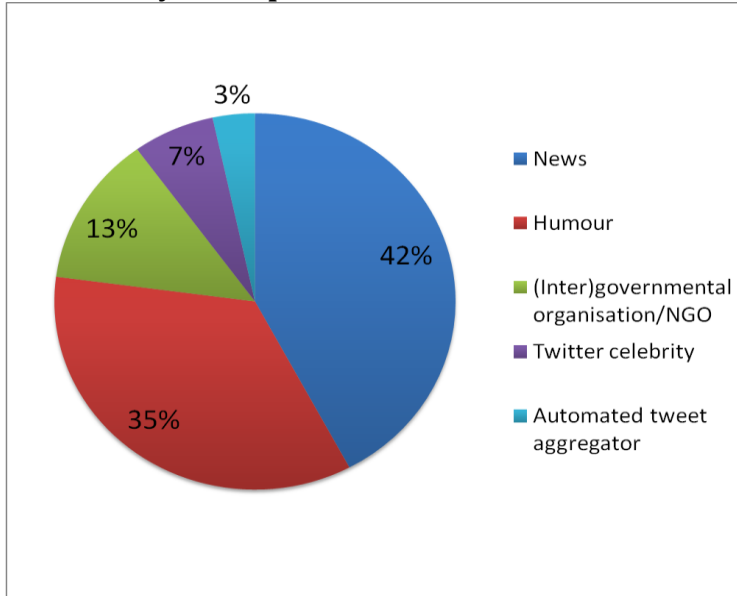


Table 1 in Appendix 6 demonstrates the features of these accounts that were used for their categorisation.

The top 10 retweeted tweets in the dataset were posted by the following account types (Table 14):

- Twitter celebrity – 5
- Celebrity – 3
- Humour – 1
- Individual – 1

Table 14. The top 10 most retweeted tweets in the dataset.

Number	Account screen name	Tweet text	Number of retweets	Lowest number of the account followers in the dataset	Account type
1	_Kyheim	Ebola jokes are funny cus it kills africans but cancer jokes aren't cus it kills white ladies . THIS TEA BOMB AF THO	62,739	2,406	Twitter celebrity
2	chrisbrown	****	45,923	13,672,027	Celebrity
3	****	****	39,749	11,512	Twitter celebrity
4	Ashton5SOS	The Ebola virus is super scary, hopefully it can be contained!	39,117	2,927,026	Celebrity
5	emmkaff	Scientists: Don't freak out about Ebola. Everyone: *Panic!*	38,633	2,175	Twitter celebrity
6	NiallOfficial	“@BandAid30: We're very happy to have your support. Thanks @NiallOfficial #BandAid30 http://www.bandaid30.com ”pleasure	37,567	20,389,619	Celebrity
7	therealbanksy	Hunger kills more than Ebola but it's not considered a significant problem since rich people can't die of it.	37,213	766,937	Humour
8	****	****	36,425	43,891	Twitter celebrity
9	jacob_binke	Ebola has been in the US for 1 day and people are already wearing masks. AIDS has been here for 55 years and people still don't use a condom	35,433	537	Individual
10	BallinInHD	This soap is \$195, it better wash Ebola, wash HIV, wash Malaria, shit... It better wash all my sins away	34,655	45,098	Twitter celebrity

**** An account has been deleted or suspended or a tweet has been deleted

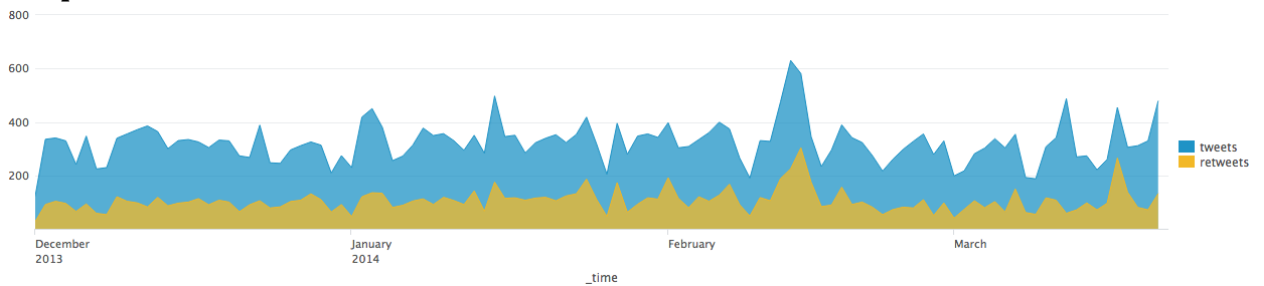
The full description of retweet features in the 5 sub-periods compared to the full study period is presented in Table 15.

Table 15. Retweet features of the 5 study sub-periods compared to the full study period.

Sub-period	Dates	Number of days (% of the number of days over the full period)	Number of tweets per sub-period (% of the number of tweets over the full period)	Number of retweets (% of the number of tweets per sub-period)	Number of unique user accounts per sub-period (daily average)	Number of unique retweeting users (% of the number of unique accounts per sub-period)	Number of unique retweeted users (% of the number of unique accounts per sub-period)	Number of retweets containing URLs over the sub-period (% of the number of retweets per sub-period)
1	1 December 2013 - 21 March 2014	111 (21.0%)	35,823 (0.1%)	11,568 (32.3%)	19,733 (178)	6,240 (31.6%)	1,917 (9.7%)	7,147 (61.8%)
2	22 March 2014 - 22 July 2014	123 (23.3%)	863,690 (1.4%)	303,647 (35.2%)	287,119 (2,334)	151,154 (52.7%)	29,798 (10.4%)	203,097 (66.9%)
3	23 July 2014 - 29 September 2014	69 (13.1%)	13,197,756 (21.8%)	4,933,633 (37.4%)	2,713,528 (39,326)	1,750,976 (64.5%)	314,265 (11.6%)	2,764,560 (56%)
4	30 September 2014 - 31 October 2014	32 (6.1%)	31,650,833 (52.3%)	17,661,398 (55.8%)	8,388,524 (262,141)	6,250,303 (74.5%)	1,069,628 (12.8%)	5,993,852 (19.0%)
5	1 November 2014 - 12 May 2015	193 (36.6%)	14,763,300 (24.4%)	6,316,984 (42.8%)	4,731,119 (24,513)	3,090,557 (65.3%)	468,616 (9.9%)	3,451,600 (54.6%)
Full period	1 December 2013 - 12 May 2015	528 (100%)	60,511,402 (100%)	29,227,230 (48.3%)	12,185,699 (23,079)	8,813,863 (72.3%)	1,565,292 (12.8%)	12,420,256 (42.5%)

In the 1st sub-period retweets accounted for 32.3% of the overall number of tweets in this sub-period. 31.6% of the unique users of this sub-period retweeted other users, whereas 9.7% of them were retweeted. 61.8% of retweets contained URLs. Figure 33 demonstrates the tweet and retweet temporal frequency distribution within the 1st sub-period.

Figure 33. Tweet and retweet temporal frequency distribution (by day) during the full period.



I categorised the top 20 retweeted accounts in the 1st sub-period into the following types:

- Twitter celebrity – 10
- Infectious disease tracking - 3
- News - 2
- Automated tweet aggregator – 1
- Inter(governmental) organisation/ NGO - 1
- Other non-verified – 1
- Celebrity – 1
- Humour – 1

Figure 34 demonstrates the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 1st sub-period.

Figure 34. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 1st sub-period.

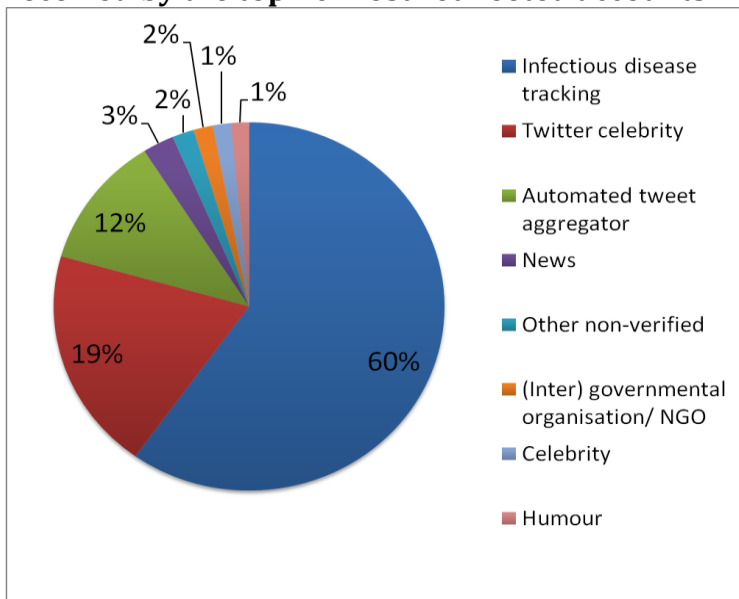


Table 2 in Appendix 6 demonstrates the features of these accounts that were used for their categorisation.

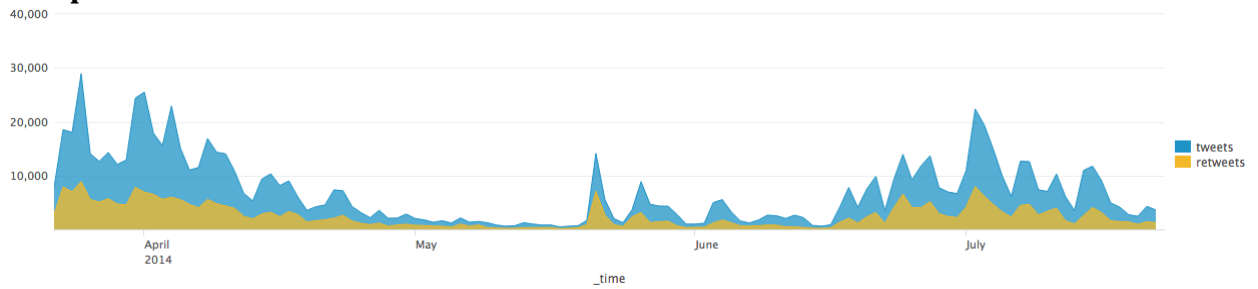
The top 10 retweeted tweets in the 1st sub-period were posted by the following account types:

- Twitter celebrity – 5
- Humour – 2
- Celebrity – 2
- News – 1

Table 1 in Appendix 7 demonstrates top 10 most retweeted tweets in the 1st sub-period and the type of accounts they were posted by.

In the 2nd sub-period retweets accounted for 35.2% of the overall number of tweets in this sub-period. 52.7% of the unique users of this sub-period retweeted other users, whereas 10.4% of them were retweeted. 66.9% of retweets contained URLs. Figure 35 demonstrates the retweet and tweet temporal frequency distribution within the 2nd sub-period.

Figure 35. Tweet and retweet temporal frequency distribution (by day) in the 2nd sub-period.



I categorised the top 20 retweeted accounts in the 2nd sub-period into the following types:

- News – 13
- Inter(governmental) organisation/ NGO – 2
- Infectious disease tracking – 2
- Twitter celebrity – 2
- Automated tweet aggregator – 1

Figure 36 demonstrates the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 2nd sub-period.

Figure 36. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 2nd sub-period.

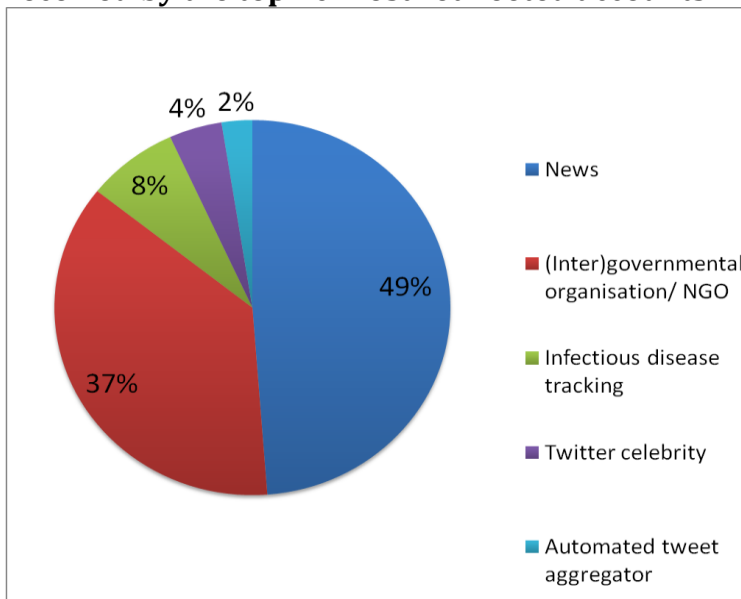


Table 3 in Appendix 6 demonstrates the features of these accounts that were used for their categorisation.

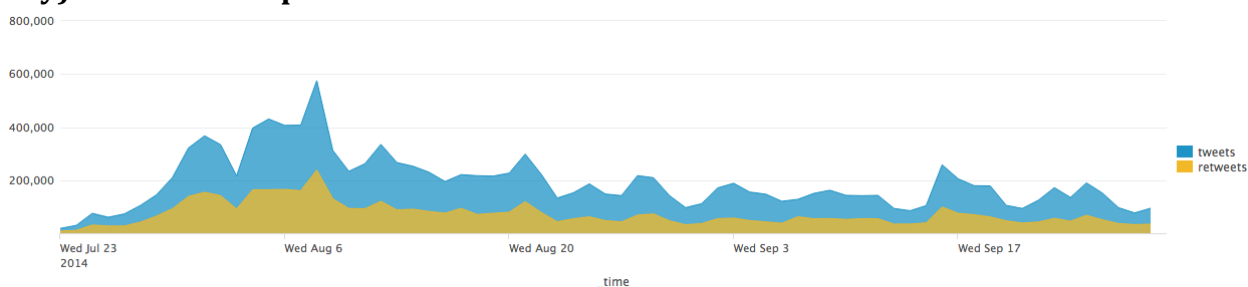
The top 10 retweeted tweets in the 2nd sub-period were posted by the following types of accounts:

- Twitter celebrity – 4
- News – 4
- Inter(governmental) organisation/ NGO – 2

Table 2 in Appendix 7 demonstrates top 10 most retweeted tweets in the 2nd sub-period and the type of accounts they were posted by.

In the 3rd sub-period retweets accounted for 37.4% of the overall number of tweets in this sub-period. 64.5% of the unique users of this sub-period retweeted other users, whereas 11.6% of them were retweeted. 56% of retweets contained URLs. Figure 37 demonstrates the retweet and tweet temporal frequency distribution within the 3rd sub-period.

Figure 37. Time chart of the tweet and retweet temporal frequency distribution (by day) in the 3rd sub-period.



I categorised the top 20 retweeted accounts in the 3rd sub-period into the following types:

- News – 14
- Twitter celebrity – 3
- Inter(governmental) organisation/ NGO – 2
- Humour – 1

Figure 38 demonstrates the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 3rd sub-period.

Figure 38. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 3rd sub-period.

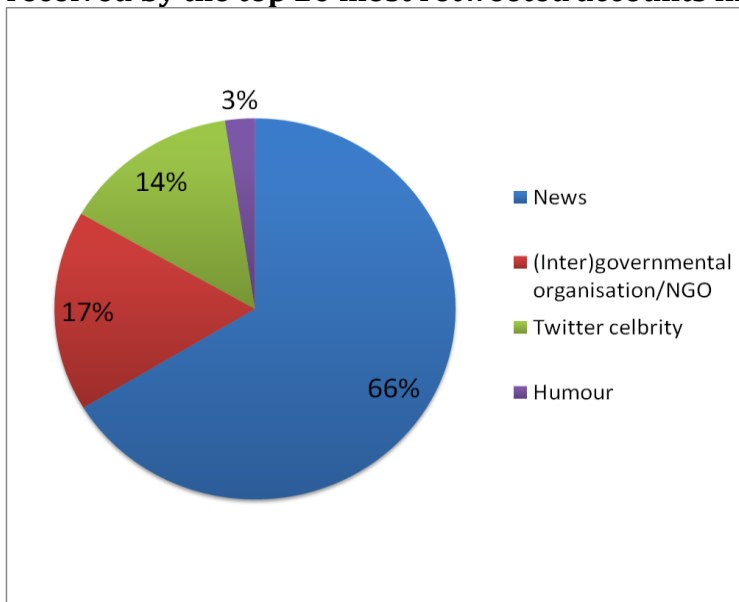


Table 4 in Appendix 6 demonstrates the features of these accounts that were used for their categorisation.

Top 10 retweeted tweets in the 3rd sub-period were posted by the following account types:

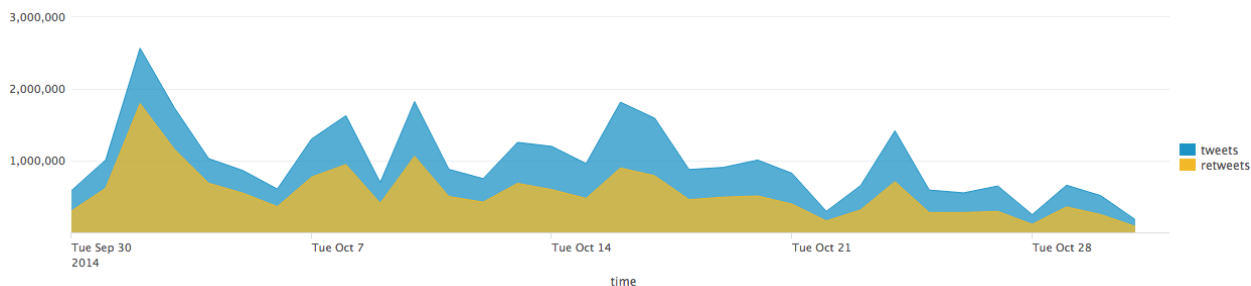
- Twitter celebrity – 7
- News – 1
- Humour – 1
- Individual – 1

Table 3 in Appendix 7 demonstrates top 10 most retweeted tweets in the 3rd sub-period and the type of accounts they were posted by.

In the 4th sub-period retweets accounted for 55.8% of the overall number of tweets in this sub-period. 74.5% of the unique users of this sub-period retweeted other users, whereas 12.8% of them were retweeted. 19% of retweets contained URLs. Figure 39

demonstrates the retweet and tweet temporal frequency distribution within the 4th sub-period.

Figure 39. Tweet and retweet temporal frequency distribution (by day) in the 4th sub-period.



I categorised the top 20 retweeted accounts in the 4th sub-period into the following types:

- Humour – 10
- News – 6
- Twitter celebrity – 4

Figure 40 demonstrates the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 4th sub-period.

Figure 40. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 4th sub-period.

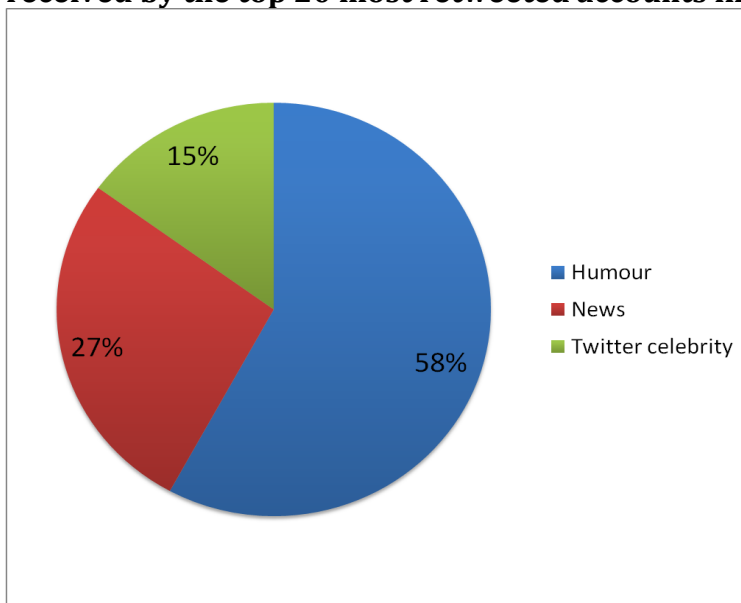


Table 5 in Appendix 6 demonstrates the features of these accounts that were used for their categorisation.

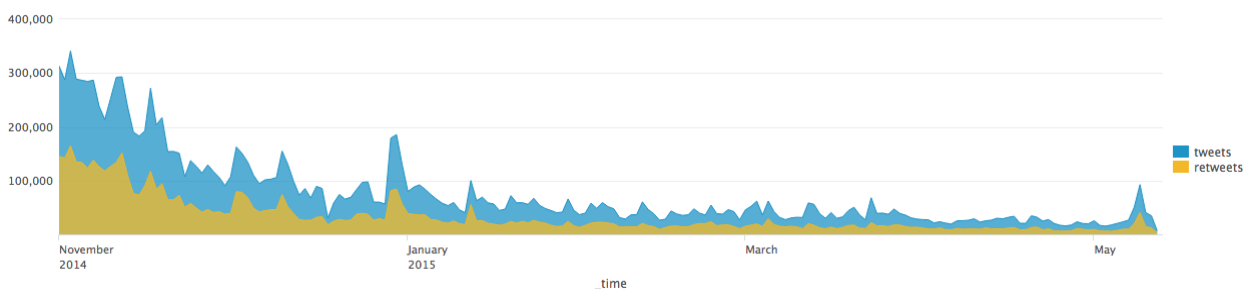
Top 10 retweeted tweets in the 4th sub-period were posted by the following types of accounts:

- Twitter celebrity – 5
- Celebrity – 3
- Humour – 1
- Individual – 1

Table 4 in Appendix 7 demonstrates top 10 most retweeted tweets in the 4th sub-period and the type of accounts they were posted by.

In the 5th sub-period retweets accounted for 42.8% of the overall number of tweets in this sub-period. 65.3% of the unique users of this sub-period retweeted other users, whereas 9.9% of them were retweeted. 54.6% of retweets contained URLs. Figure 41 demonstrates the retweet and tweet temporal frequency distribution within the 5th sub-period.

Figure 41. Tweet and retweet temporal frequency distribution (by day) in the 5th sub-period.



I categorised the top 20 retweeted accounts in the 5th sub-period into the following types:

- News - 6
- Inter(governmental) organisation/ NGO – 5
- Celebrity – 4
- Humour -3
- Automated tweet aggregator – 1

- Twitter celebrity – 1

Figure 42 demonstrates the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 5th sub-period.

Figure 42. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts in the 5th sub-period.

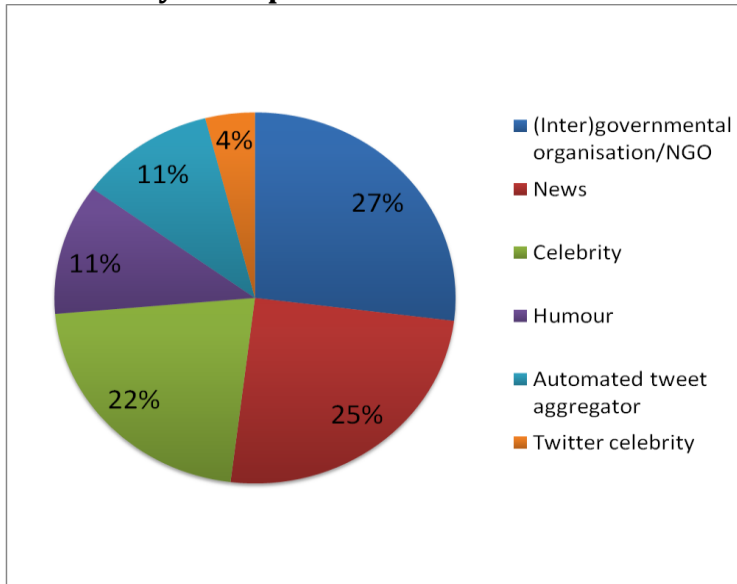


Table 6 in Appendix 6 demonstrates the features of these accounts that were used for their categorisation.

The top 10 retweeted tweets in the 5th sub-period were posted by the following account types:

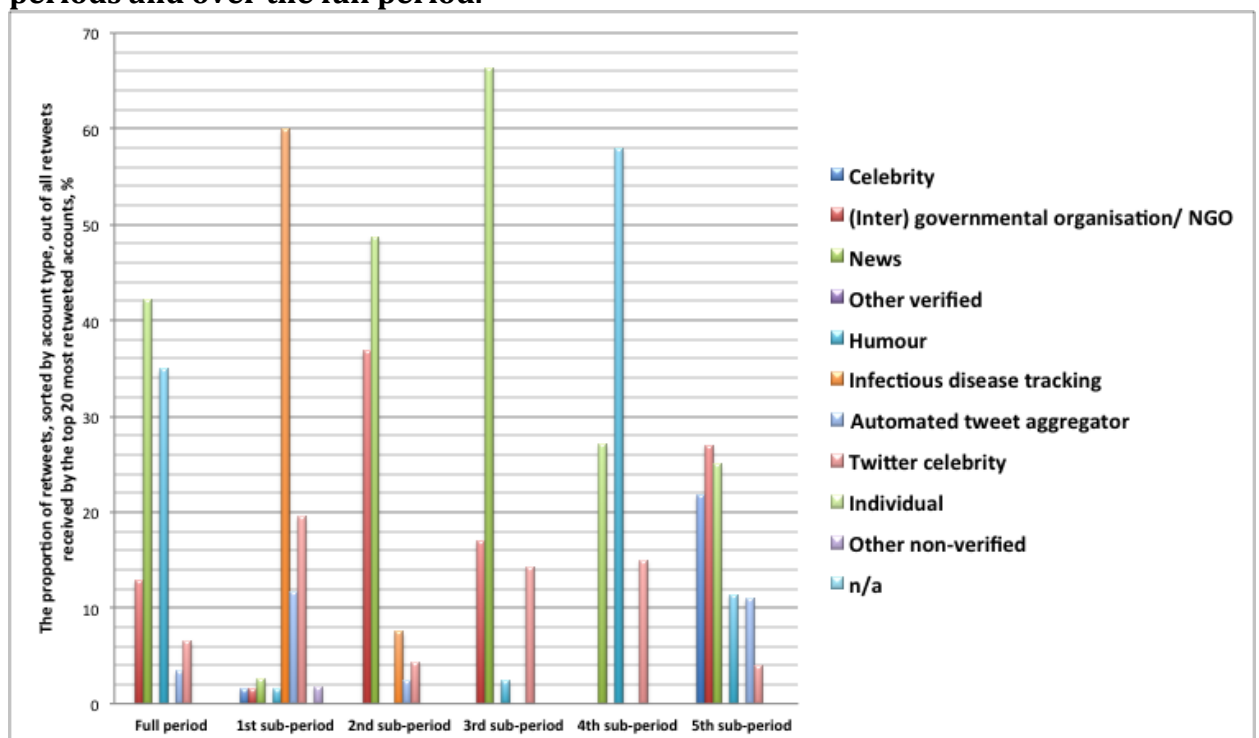
- Celebrity – 5
- Humour – 3
- Twitter celebrity – 1
- Other verified – 1

Table 5 in Appendix 7 demonstrates top 10 most retweeted tweets in the 5th sub-period and the type of accounts they were posted by.

Figure 43 summarises the findings above, presenting the proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts within each sub-period and over the full period. Overall, the most retweeted types of

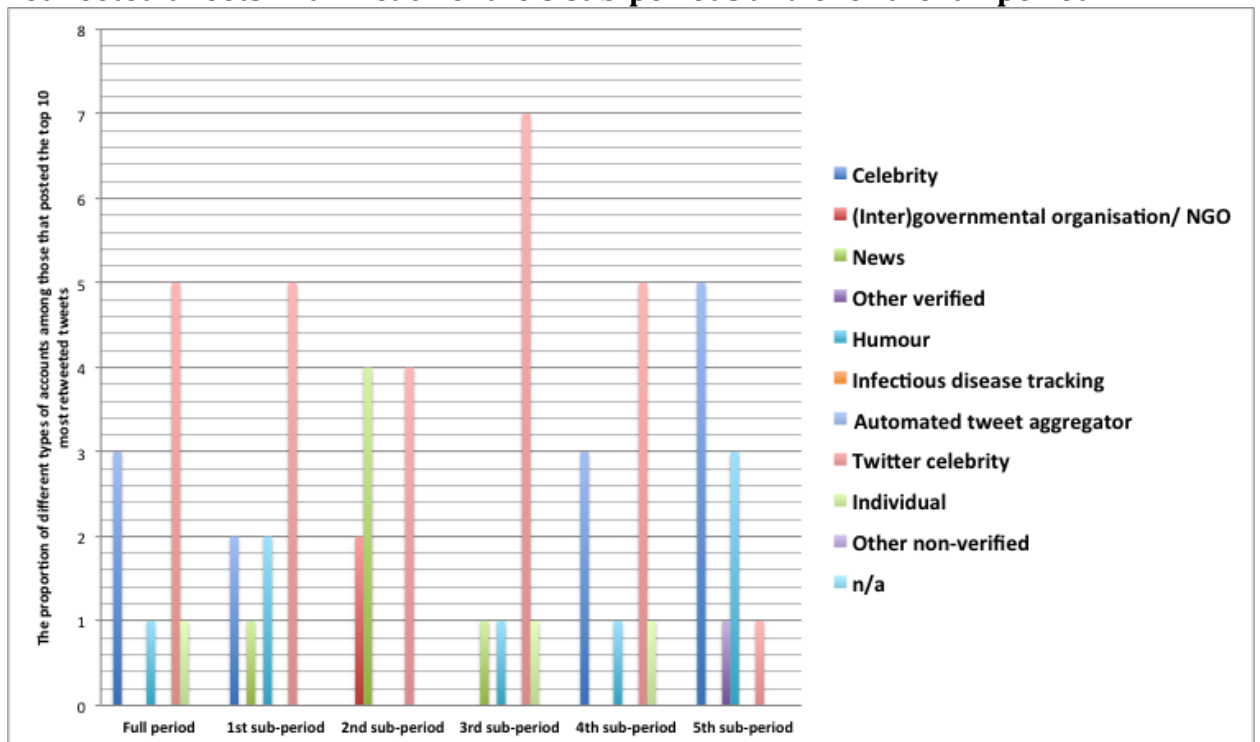
accounts were those of news organisations, humour accounts, Twitter celebrities, and (inter)governmental organisations/NGOs. It is evident that in the 1st baseline sub-period the diversity of the types of most retweeted accounts was greater than in any other sub-period. The most retweeted accounts in the 1st sub-period were infectious disease tracking accounts, whereas during the 2nd and the 3rd sub-periods they were accounts of news organisations. During the 4th sub-period humour accounts received the most retweets, with news accounts still being highly retweeted. During the last, 5th sub-period, accounts of (inter)governmental organisations/NGOs were the most retweeted accounts, followed by accounts of news organisations and celebrity accounts.

Figure 43. The proportion of retweets, sorted by account type, out of all retweets received by the top 20 most retweeted accounts within each of the 5 study sub-periods and over the full period.



As Figure 44 demonstrates, the most retweeted tweets in the dataset were posted by the accounts of Twitter celebrities, celebrities, humour, and news accounts. The most retweeted tweets of the first 4 sub-periods were mainly posted by Twitter celebrities and the most retweeted tweets of the 5th sub-period were posted mainly by celebrities.

Figure 44. The proportion of different types of accounts that posted the most retweeted tweets within each of the 5 sub-periods and over the full period.



5.5. Mentions

Figure 45 demonstrates the tweet, “any mention”, mention⁴⁷ and reply temporal frequency distribution over the full study period. Table 16 demonstrates the number of tweets containing “any mentions”, mentions, and replies and their prevalence among the total number of tweets in the dataset. 61% of the tweets contained “any mentions”, 12.8% contained mentions, and 6.6% contained replies. The highest prevalence of mentions and replies was observed in the baseline sub-period (25.2% and 14.8% respectively). The prevalence of mentions was 13%, 13.6%, 12.2%, and 13.5% and the prevalence of replies was 4.9%, 4.8%, 7.1%, and 7.1% during the 2nd, 3rd, 4th, and 5th sub-periods respectively.

⁴⁷ Mentions without retweets

Figure 45. Temporal frequency distribution (by day) of tweets, “any mentions”, mentions and replies over the full study period.

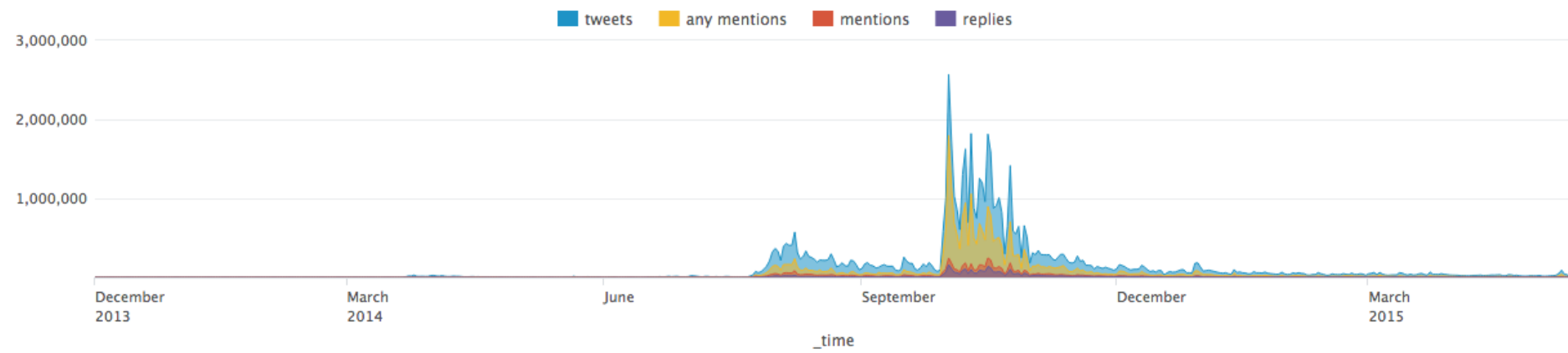


Table 16. The number of tweets containing “any mentions”, mentions, and replies and their prevalence of the total number of tweets in the dataset.

Sub-period	Dates	Number of days (% of the number of days over the full period)	Number of tweets per sub-period (daily average)	Number of tweets containing "any mentions" (% of the number of tweets per sub-period)	Number of tweets containing mentions (% of the number of tweets per sub-period)	Number of tweets containing replies (% of the number of tweets per sub-period)
1	1 December 2013 - 21 March 2014	111 (21.0%)	35,823 (323)	12,119 (33.8%)	9,032 (25.2%)	5,304 (14.8%)
2	22 March 2014 - 22 July 2014	123 (23.3%)	863,690 (7,022)	317,152 (36.7%)	112,754 (13.0%)	42,135 (4.9%)
3	23 July 2014 - 29 September 2014	69 (13.1%)	13,197,756 (191,272)	6,723,764 (51.0%)	1,790,131 (13.6%)	634,186 (4.8%)
4	30 September 2014 - 31 October 2014	32 (6.1%)	31,650,833 (989,089)	21,524,062 (68.0%)	3,862,664 (12.2%)	2,243,249 (7.1%)
5	1 November 2014 - 12 May 2015	193 (36.6%)	14,763,300 (76,494)	8,310,560 (56.3%)	1,993,576 (13.5%)	1,050,012 (7.1%)
Full period	1 December 2013 - 12 May 2015	528 (100%)	60,511,402 (114,605)	36,887,657 (61.0%)	7,768,157 (12.8%)	3,974,886 (6.6%)

Table 17. The number of unique mentioning and mentioned, replying and replied user accounts.

Sub-period	Dates	Number of days (% of the number of days over the full period)	Number of unique user accounts per period (daily average)	Number of unique "any mentioning" accounts (% of the number of unique accounts per sub-period)	Number of unique "any mentioned" accounts (% of the number of unique accounts per sub-period)	Number of unique mentioning accounts (% of the number of unique accounts per sub-period)	Number of unique mentioned accounts (% of the number of unique accounts per sub-period)	Number of unique replying accounts (% of the number of unique accounts per sub-period)	Number of unique accounts received a reply (% of the number of unique accounts per sub-period)
1	1 December 2013 - 21 March 2014	111 (21.0%)	19,733 (178)	7,479 (37.9%)	7,055 (35.8%)	6,061 (30.7%)	6,990 (35.4%)	4,301 (21.8%)	4,581 (23.0%)
2	22 March 2014 - 22 July 2014	123 (23.3%)	287,119 (2,334)	159,081 (55.4%)	53,151 (18.5%)	57,241 (20.0%)	39,622 (13.8%)	25,948 (9.0%)	23,101 (8.1%)
3	23 July 2014 - 29 September 2014	69 (13.1%)	2,713,528 (39,326)	2,110,879 (77.8%)	587,647 (21.7%)	586,713 (21.6%)	395,885 (14.6%)	311,095 (11.5%)	250,211 (9.2%)
4	30 September 2014 - 31 October 2014	32 (6.1%)	8,388,524 (262,141)	7,075,149 (84.3%)	1,960,957 (23.4%)	1,672,949 (20.0%)	1,259,918 (15.0%)	1,150,072 (13.7%)	833,298 (9.9%)
5	1 November 2014 - 12 May 2015	193 (36.6%)	4,731,119 (24,513)	3,786,704 (80.0%)	1,144,497 (24.2%)	933,872 (19.7%)	821,076 (17.4%)	612,945 (13.0%)	549,694 (11.6%)
Full period	1 December 2013 - 12 May 2015	528 (100%)	12,185,699 (23,079)	10,278,157 (84.4%)	3,094,532 (25.4%)	2,702,493 (22.2%)	2,132,632 (17.5%)	1,845,256 (15.1%)	1,442,048 (11.8%)

Table 17 demonstrates the number of unique mentioning and mentioned, replying and replied user accounts. On average 22.2% of the unique user accounts in the dataset mentioned other user accounts, 17.5% were mentioned, 15.1% were replying to other users, and 11.8% received replies. The highest mentioning and replying activity was observed in the 1st baseline sub-period. The lowest proportion of replying and receiving replies accounts was observed during the 2nd sub-period (9.0% for replying accounts and 8.1% for receiving replies accounts), which was increased in the following sub-periods. The 2nd sub-period had also the lowest proportion of mentioned accounts, whereas the 5th sub-period had the lowest proportion of mentioning accounts (19.7%).

Figures 46, 47, 48, 49, 50 and 51 below present the pie charts of the proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the full study period and each of the 5 sub-periods.

Figure 46. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the dataset.

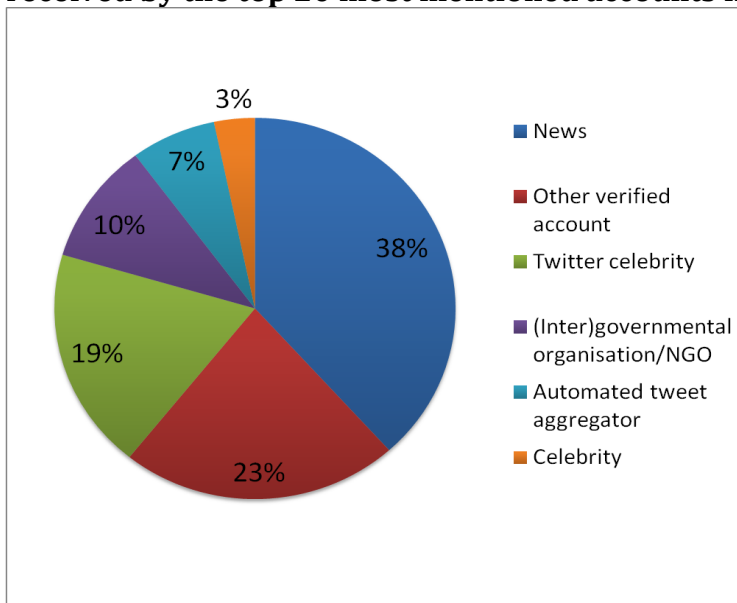
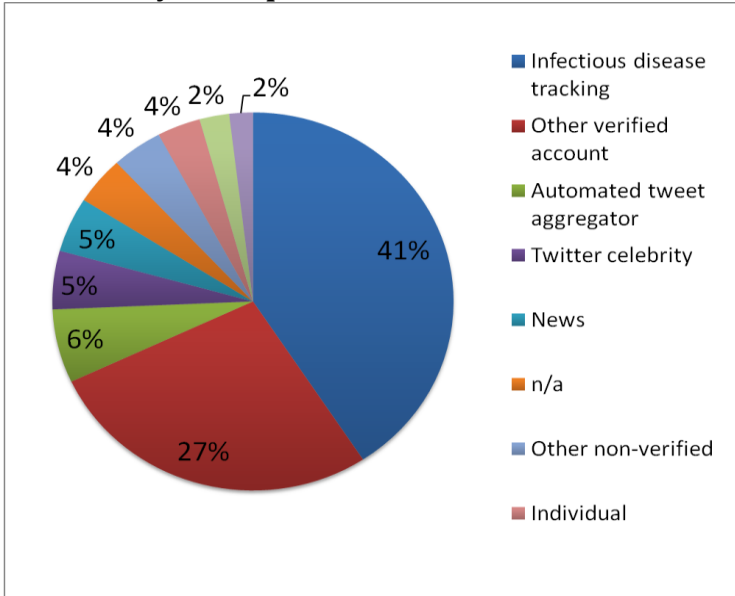
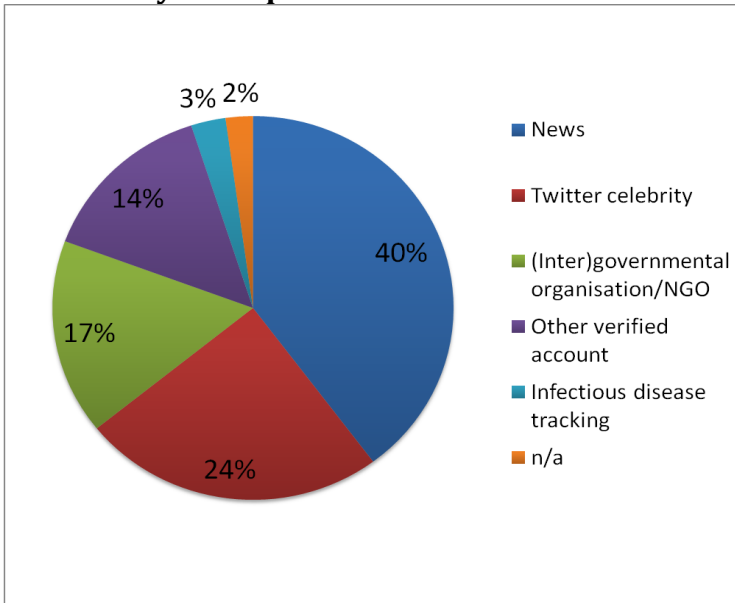


Figure 47. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 1st sub-period.



n/a means that it was not possible to identify the account type, as some of the information was missing.

Figure 48. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 2nd sub-period.



n/a means that it was not possible to identify the account type, as some of the information was missing.

Figure 49. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 3rd sub-period.

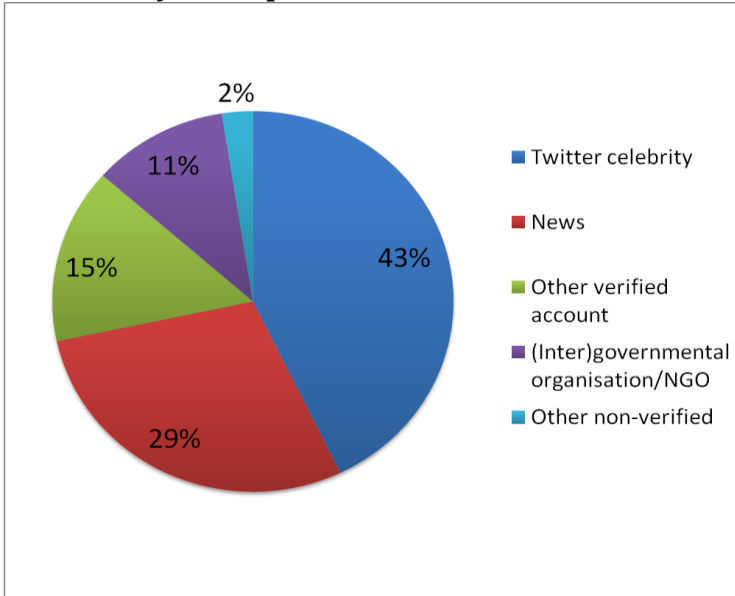


Figure 50. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 4th sub-period.

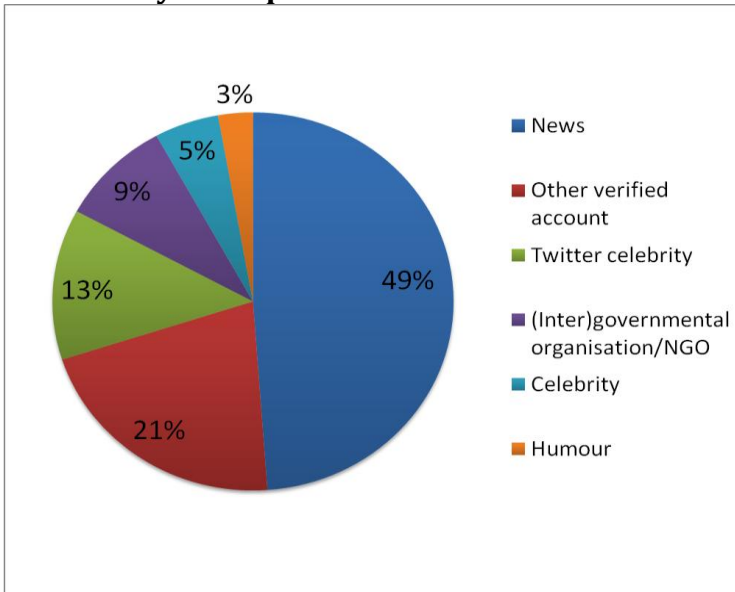
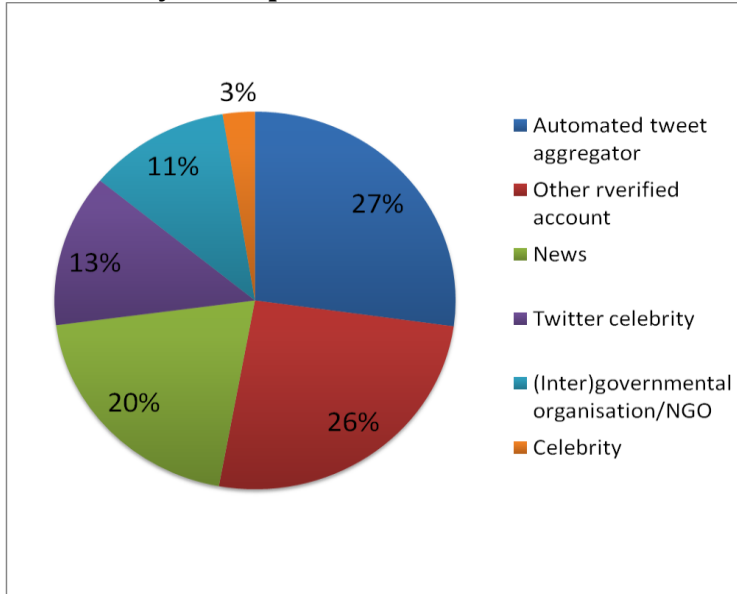
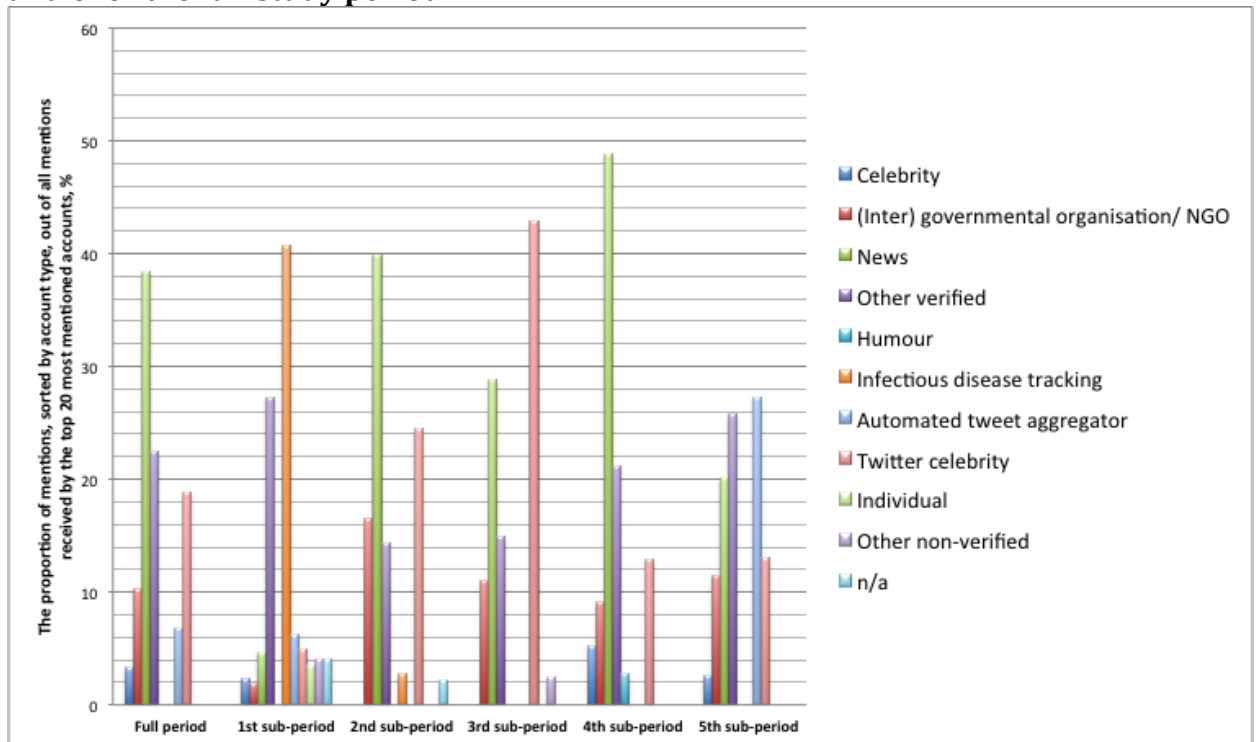


Figure 51. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts in the 5th sub-period.



Tables 1, 2, 3, 4, 5 and 6 in Appendix 8 present the features of these accounts that I used for their categorisation. As can be seen from Figure 52, the most mentioned types of accounts were those of news organisation, other verified accounts (such as YouTube), Twitter celebrities, and (inter)governmental organisations/NGOs. Infectious disease tracking accounts were highly mentioned in the 1st study sub-period. Twitter accounts of (inter)governmental organisations/NGOs received the highest number of mentions in the 2nd sub-period, Twitter celebrities in the 3rd sub-period, news organisations in the 4th sub-period, and automated tweet aggregators in the 5th sub-period. .

Figure 52. The proportion of mentions, sorted by account type, out of all mentions received by the top 20 most mentioned accounts within each of the 5 sub-periods and over the full study period.

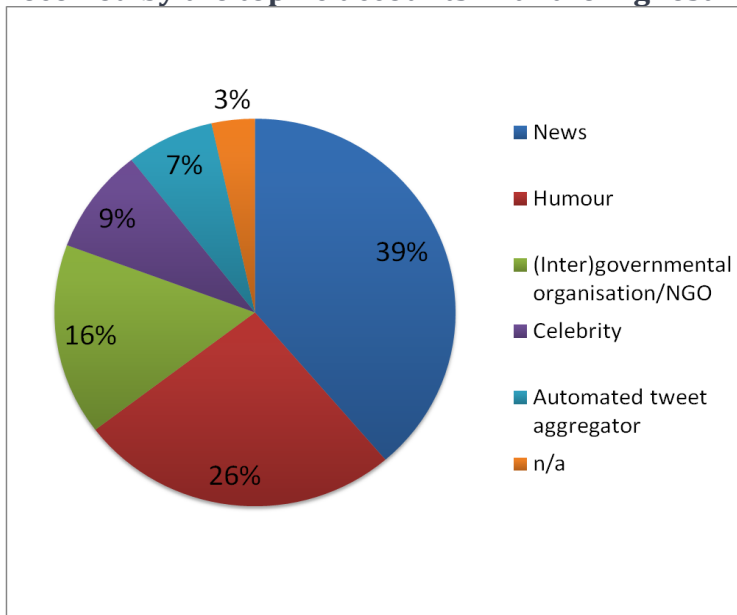


n/a means that it was not possible to identify the account type, as some of the information was missing.

5.6. Replies

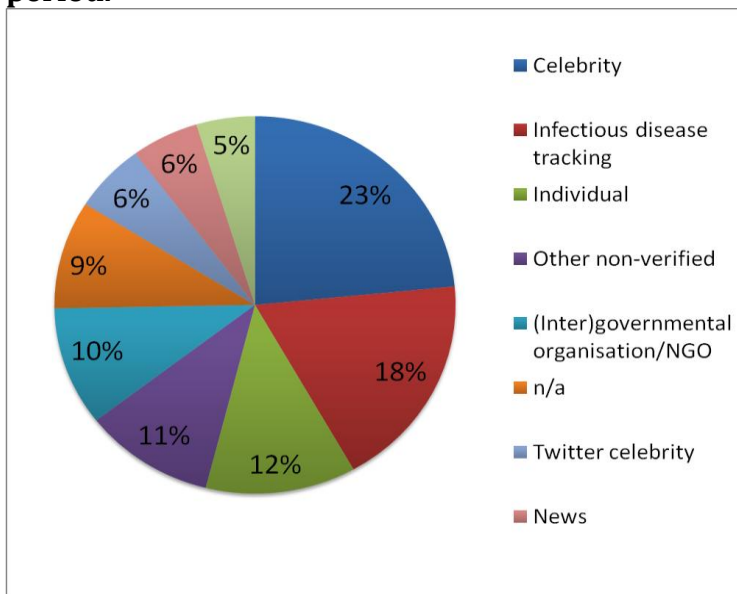
Figures 53, 54, 55, 56, 57, and 58 below present the pie charts of the proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the full period and in each of the 5 sub-periods.

Figure 53. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the dataset.



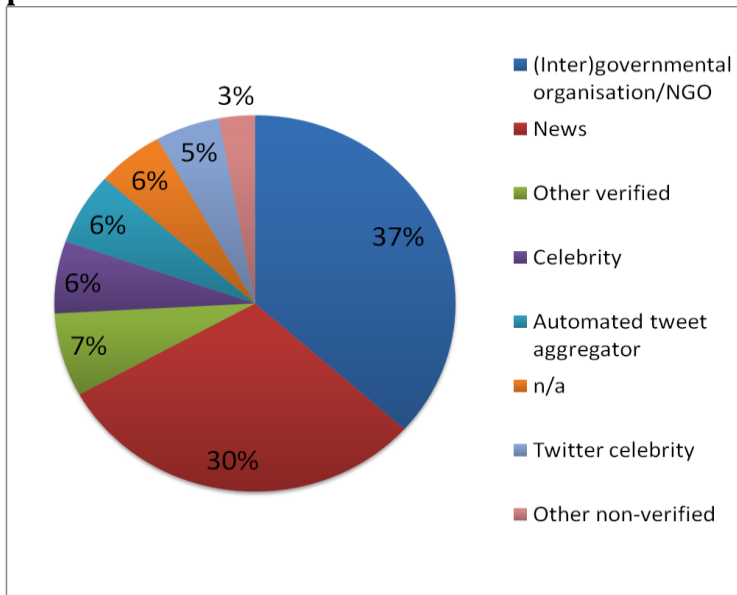
n/a means that it was not possible to identify the account type, as some of the information was missing.

Figure 54. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 1st sub-period.



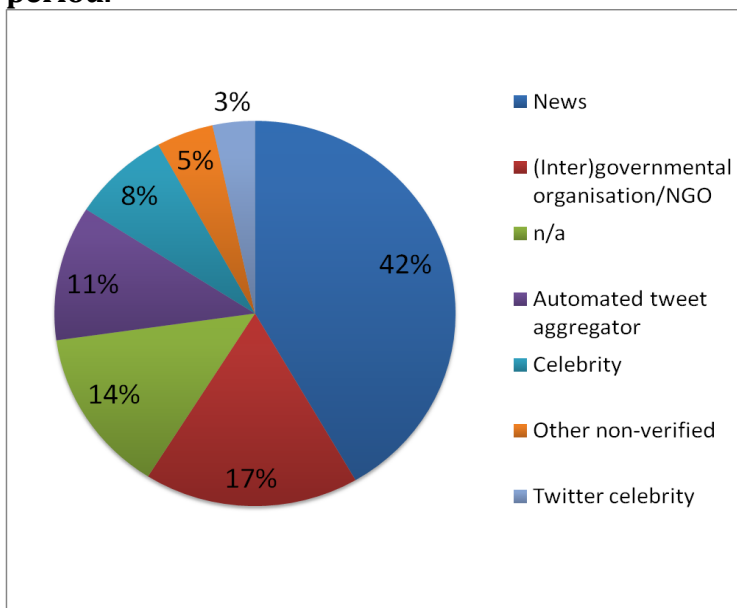
n/a means that it was not possible to identify the account type, as some of the information was missing.

Figure 55. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 2nd sub-period.



n/a means that it was not possible to identify the account type, as some of the information was missing.

Figure 56. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 3rd sub-period.



n/a means that it was not possible to identify the account type, as some of the information was missing.

Figure 57. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 4th sub-period.

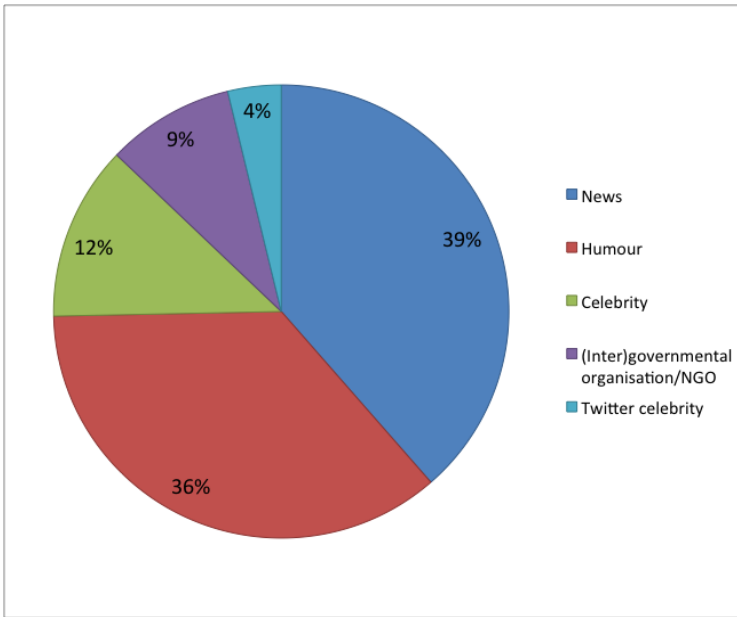
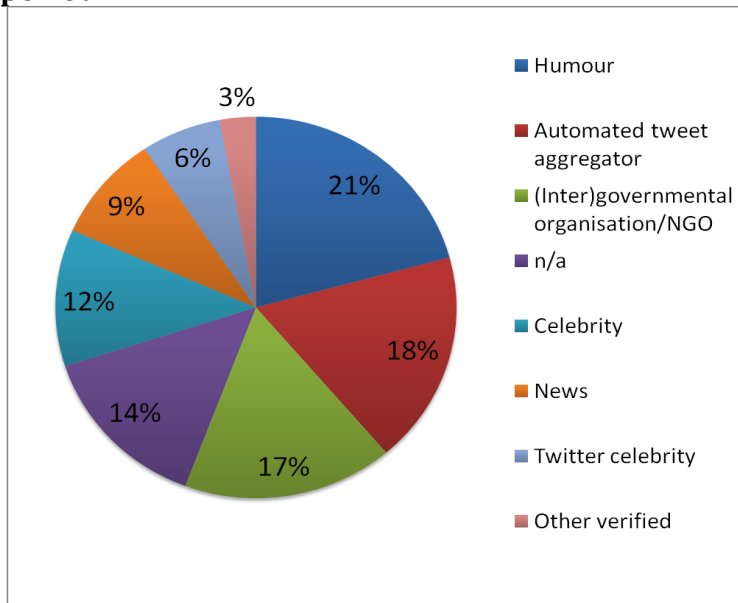


Figure 58. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies in the 5th sub-period.

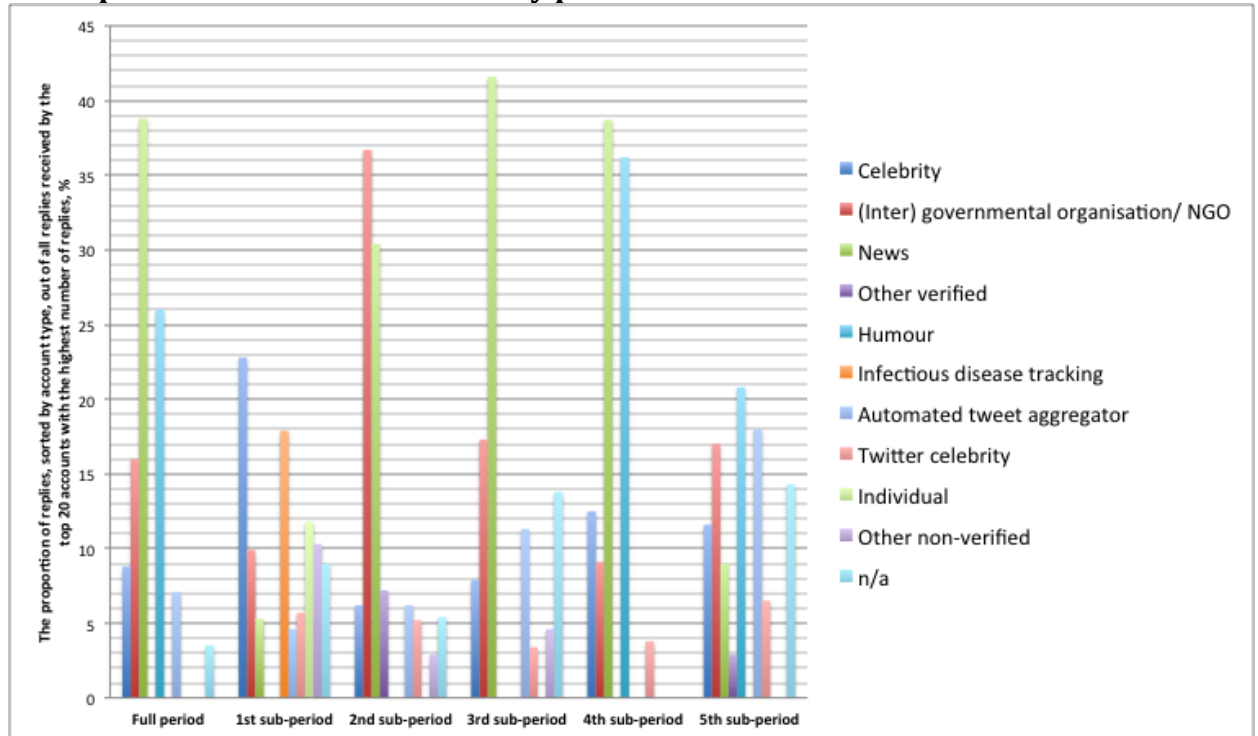


n/a means that it was not possible to identify the account type, as some of the information was missing.

Tables 1, 2, 3, 4, 5 and 6 in Appendix 9 present the features of these accounts that I used for their categorisation. As can be seen from Figure 59, accounts that received the highest number of replies were those of news organisations, humour, (inter)governmental organisations/NGOs, celebrities, and automated tweet aggregators. Accounts of celebrities received the highest number of replies in the 1st sub-period, (inter)governmental

organisations/NGOs in the 2nd sub-period, news organisations in the 3rd and 4th sub-periods, and humour accounts in the 5th sub-period.

Figure 59. The proportion of replies, sorted by account type, out of all replies received by the top 20 accounts with the highest number of replies within each of the 5 sub-periods and over the full study period.



n/a means that it was not possible to identify the account type, as some of the information was missing.

5.7. Retweets, mentions, and replies of (inter)governmental organisation/NGO accounts

Table 18 demonstrates the organisations owning the (inter)governmental organisation/NGO type of accounts that were identified among the top 20 most retweeted, mentioned, and the top 20 that received the highest number of replies.

Overall, the WHO account was the third and UNICEF was the seventh most retweeted accounts in the dataset. The WHO account was the fifth and CDC account was the eighth most mentioned accounts in the dataset. None of the top 10 most retweeted tweets was posted by any of the (inter)governmental organisation/NGO accounts. Moreover, the WHO, CDC and White House were among the top 20 accounts that received the highest number of replies, with the WHO account being the second overall.

In the 1st sub-period, the account of the Australian Government Department of Health was among the top 20 most retweeted accounts, whereas the WHO account was one of the top 20 mentioned accounts and the second account that received the highest number of replies.

In the 2nd sub-period the WHO account was the most retweeted account, with the United Nations (UN) account being one of the top 20 retweeted accounts. 2 of the top 10 retweeted tweets were posted by these accounts in the 2nd sub-period. Moreover, the WHO account was the second most mentioned account in this period and the account that received the highest number of replies.

In the 3rd sub-period the WHO account was the most retweeted account with UNICEF being one of the top 20 retweeted accounts. The WHO account was the third most mentioned account and the CDC account was among the top 20 mentioned accounts. The WHO account received the highest number of replies in this sub-period, with the CDC account being one of the top 20 with the highest number of replies.

In the 4th sub-period none of the (Inter)governmental organisation/NGO accounts were among the top 20 retweeted accounts or posted any of the top 10 retweeted tweets. However, CDC account was the fourth most mentioned account and the WHO account was among the top 20 most mentioned accounts. The accounts of CDC and White house were among the top 20 accounts that received the highest number of replies.

In the 5th sub-period UNICEF account was the second and the WHO account was the fifth most retweeted accounts. The UN, MSF (Doctors Without Borders) Spain and UN Mission for Ebola Emergency Response (UNMEER) were among the top 20 retweeted accounts. The WHO, UN, UNMEER, and UNICEF accounts were among the top 20 mentioned accounts and the top 20 accounts with the highest number of replies. Moreover, the WHO was the third account with the highest number of replies in this sub-period.

Table 18. Organisations owning (inter)governmental organisation/NGO accounts among the top 20 retweeted, mentioned, with the highest number of replies, or the top 10 that posted the most retweeted tweets.

	Full period	1st sub-period	2nd sub-period	3rd sub-period	4th sub-period	5th sub-period
Top 20 retweeted accounts	WHO UNICEF	Australian Government Department of Health	WHO UN	WHO UNICEF	none	UNICEF WHO UN MSF Spain UNMEER
Top 10 most retweeted tweets	none	none	WHO UN	none	none	none
Top 20 mentioned accounts	WHO CDC	WHO	WHO	WHO CDC	CDC WHO	WHO UN UNMEER UNICEF
Top 20 accounts that received the highest number of replies	WHO CDC White House	WHO	WHO	WHO CDC	CDC White House	WHO UN UNMEER UNICEF

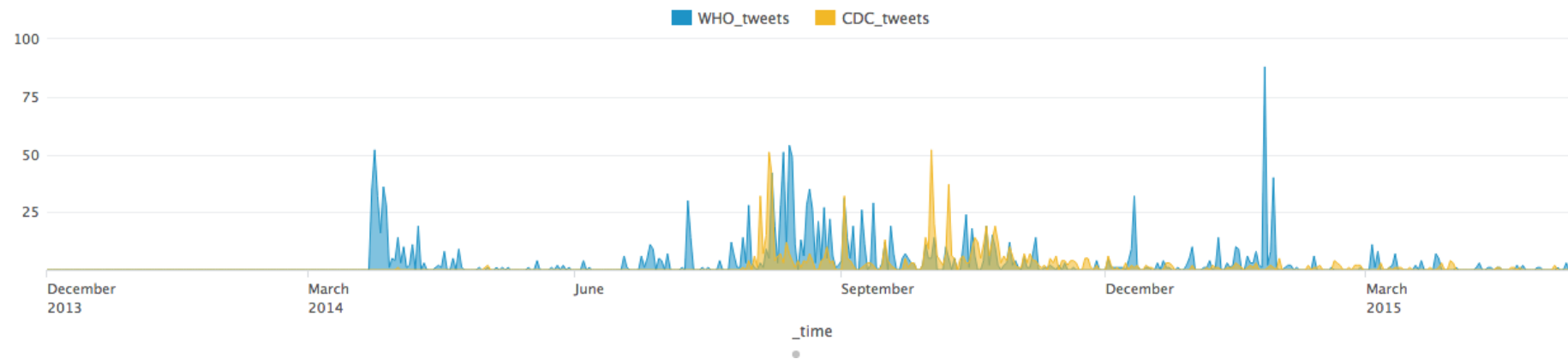
5.8. WHO and CDC Twitter Ebola activity

Table 19 demonstrates the overview of the Twitter activity of the WHO and CDC during the Ebola outbreak.

Table 19. Description of the Twitter activity of the WHO and CDC during the Ebola outbreak.

	Total number of tweets	Total number of retweets	Number of unique retweeting users	Number of mentions	Number of unique mentioning users	Number of replies	Number of unique replying users	The lowest number of followers in the dataset	The highest number of followers in the dataset
WHO	1,798	166,463	71,932	202,852	96,610	11,833	5,772	1,173,156	2,351,184
CDC	865	49,717	27,976	86,940	48,869	7,195	4,796	292,595	520,508

Figure 60. Temporal WHO and CDC tweet frequency distribution.



5.8.1. WHO and CDC Ebola tweets and their content

Figure 60 presents the temporal WHO and CDC tweet frequency distribution. Over the study period WHO posted 1,798 tweets. WHO Ebola Twitter activity started on 23 March 2014 (at the beginning of the second sub-period). The results of the content analysis of the WHO tweets presented in Table 20 demonstrate that 29.9% of the tweets belong to the “Responses and Twitter chats” category, 28.7% to the “Statement, announcement, and report”, 13.8% to the “Tweeting from official events”, 9.8% to the “Press/news release, scientific publication, press conference, interview”, 8.6% to the “Public health messages for the general public”, 7.4% to the “Personal story, feature story”, 1% to the “Retweets”, and 0.9% to the “Public health messages and information for health workers”. Most of the tweets within the 2nd and the 3rd sub-periods were “Responses and Twitter chats”, 40.1% and 37.6% respectively, whereas during the 4th sub-period most of the tweets, 33.3%, were of a “Press/news release, scientific publication, press conference, interview” category, and 33.7% of the tweets in the 5th sub-period were “Tweets from official events”.

The CDC Twitter account posted 865 tweets over the full study period. CDC Ebola Twitter activity started on 1 April 2014. The results of the content analysis of the CDC tweets presented in Table 21 demonstrate that 32.8% of the tweets belonged to the “Responses and Twitter chats” category, 23.1% were “Retweets”, 13.2% were the “Personal story, feature story”, 12.8% were the “Public health messages for the general public”, 11.8% were the “Statement, announcement, report”, 3.8% were the “Press/news release, scientific publication, press conference, interview”, and 3.2% were the “Public health messages for health workers”. The majority of the tweets in the 3rd and 4th sub-periods (34.8% and 45.4% of total number of tweets respectively) were “Responses and Twitter chats”. 31.4% of the tweets in the 5th sub-period were “Retweets”.

Table 20. Categorisation of the WHO tweets.

Sub-period	Number of days (% of the number of days over the full period)	Statement, announcement, report (% of the total number of tweets per sub-period)	Press/news release, scientific publication, press conference, interview (% of the total number of tweets per sub-period)	Retweets (% of the total number of tweets per sub-period)	Tweeting from official events (% of the total number of tweets per sub-period)	Personal story, feature story (% of the total number of tweets per sub-period)	Responses and Twitter chats (% of the total number of tweets per sub-period)	Public health messages for the general public (% of the total number of tweets per sub-period)	Public health messages for health workers (% of the total number of tweets per sub-period)	Total (% of the total number of tweets)
2	123	147	18	5	4	10	171	67	5	427
	25%	34.4%	4.2%	1.2%	0.9%	2.3%	40.1%	15.7%	1.2%	23.8%
3	69	241	60	1	73	32	290	67	8	772
	16.6%	31.2%	7.8%	0.1%	9.5%	4.2%	37.6%	8.7%	1.0%	42.9%
4	32	36	60	3	30	17	27	7	0	180
	7.6%	19.9%	33.3%	1.8%	16.4%	9.4%	15.2%	4.1%	0	10%
5	193	92	38	9	141	74	49	13	3	419
	46.3%	22.0%	9.0%	2.2%	33.7%	17.7%	11.6%	3.2%	0.7%	23.3%
Total	417	516	176	18	248	133	537	154	16	1,798
		28.7%	9.8%	1%	13.8%	7.4%	29.9%	8.6%	0.9%	100%

Table 21. Categorisation of the CDC tweets.

Sub-period	Number of days (% of the number of days over the full period)	Statement, announcement, report (% of the total number of tweets per sub-period)	Press/news release, scientific publication, press conference, interview (% of the total number of tweets per sub-period)	Retweets (% of the total number of tweets per sub-period)	Personal story, feature story (% of the total number of tweets per sub-period)	Responses and Twitter chats (% of the total number of tweets per sub-period)	Public health messages for the general public (% of the total number of tweets per sub-period)	Public health messages for health workers (% of the total number of tweets per sub-period)	Total (% of the total number of tweets)
2	123	0	0	2	2	0	0	0	4
	25%	0	0	50.0%	50.0%	0	0	0	0.5%
3	69	32	31	59	68	120	29	6	345
	16.6%	9.2%	8.9%	17%	19.8%	34.8%	8.4%	1.7%	40.8%
4	32	31	0	80	8	149	49	11	328
	7.6%	9.5%	0	24.4%	2.4%	45.4%	14.9%	3.4%	37.3%
5	193	39	2	59	36	10	32	10	188
	46.3%	20.8%	1.1%	31.4%	19.2%	5.3%	17.0%	5.3%	21.4%
Total	417	102	33	200	114	284	111	28	865
		11.8%	3.8%	23.1%	13.2%	32.8%	12.8%	3.2%	100.0%

WHO and CDC tweets on peak days throughout the dataset

The highest number of tweets per day (40 and more) was posted by the WHO on the following dates: 24 March 2014 (52 tweets), 8 August 2014 (42 tweets), 12 August 2014 (51 tweets), 14 August 2014 (54 tweets), 25 January 2015 (88 tweets), and 28 January 2015 (40 tweets). Table 22 demonstrates the distribution of the WHO tweets by category on the peak days. On 24 March 2014, most of the tweets (69.2%) consisted of “Responses and Twitter chats” tweets, as well as “Statements” (15.4%) and “Public health messages for the general public” (13.5%). The main topic of this day was the announcement of the Ebola outbreak figures in Guinea and provision of the public health information to the general public about the ways of EVD transmission. On 8 August 2014, most of the tweets (47.6%) consisted of “Responses and Twitter chats” tweets, as well as “Statements” (31%), “Tweeting from official events” (19%), and “Public health messages for the general public” (2.4%). On this day the first meeting of the International Health Regulations Emergency Committee on the 2014 Ebola Outbreak in West Africa happened, and the Ebola outbreak was declared a Public Health Emergency of International Concern. Therefore, this meeting was covered live on Twitter and the recommendations for the member countries were announced. On 12 August 2014 the WHO meeting on ethical considerations for the use of the experimental medicines for Ebola was held and was covered live on Twitter (25.5% of the tweets on this day). The press conference about this was organised and also tweeted live (15.7%). The statements of this meeting were released (16.7%). On 14 August 2014 WHO continued the communication about Ebola on Twitter, delivering information about EVD through statements (16.7%), public health messages for the general public (16.7%) and through direct responses (64.8%). On 25 January 2015, WHO Executive board session on Ebola started and was covered on Twitter via tweeting from the event (89.9%) and a media briefing (10.2%). On 28 January 2015 WHO organised a Twitter chat with the

hashtag #AskRebecca, providing the Twitter users with the opportunity to ask Rebecca, a nurse and an Ebola survivor from Sierra Leone, questions.

The highest number of CDC tweets (more than 30) was posted on the following 5 days: 4 August 2014 (32 tweets), 7 August 2014 (51 tweets), 8 August 2014 (41 tweets), 2 October 2014 (52 tweets), and 8 October 2014 (37 tweets). Table 23 demonstrates the distribution of the CDC tweets by type on these peak days. On 4 August 2014 CDC organised a Twitter chat allowing anyone on Twitter to ask CDC experts questions about Ebola. 93.8% of the tweets on this day were a part of this chat. On 7 August 2014 CDC was preparing for their next Twitter chat on 8 August 2014, sending direct invitations to Twitter users (90.2% of tweets on this day) to join the chat. On 8 August 2014 CDC organised a Twitter chat for clinicians, providing the information on what the US hospitals need to know about Ebola and infection control (87.8% of the tweets on this day). On 2 October 2014, the third Twitter chat was organised providing the information on what the US public needs to know on the first diagnosed Ebola case in the USA (94.3% of tweets on this day). The fourth CDC chat on Ebola was organised on 8 October 2014 (81.1% of the tweets on this day).

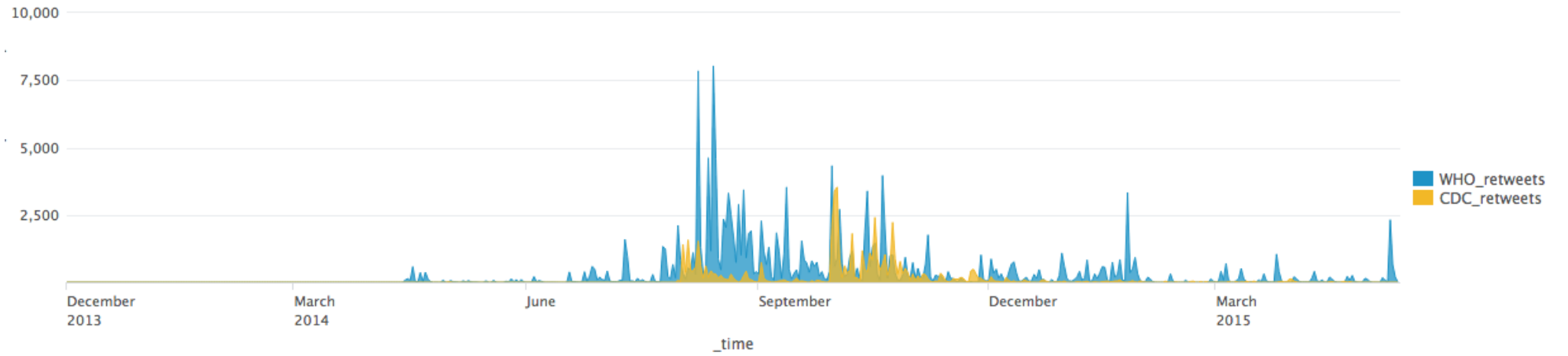
Table 22. Distribution of the WHO tweets by type on peak days.

Date	Statement, announcement, report (% of the total number of tweets per period)	Press/news release, scientific publication, press conference, interview	Retweets	Tweeting from official events	Personal story, feature story	Responses and Twitter chats	Public health messages for the general public	Public health messages for health workers	Total
24 March, 2014	8	0	1	0	0	36	7	0	52
% of the number of tweets on this day	15.4%		1.9%			69.2%	13.5%		
8 August, 2014	13	0	0	8	0	20	1	0	42
% of the number of tweets on this day	31%			19.0%		47.6%	2.4%		
12 August, 2014	14	8	0	13	0	15	1	0	51
% of the number of tweets on this day	27.5%	15.7%		25.5%		29.4%	2.0%		
14 August, 2014	9	0	0	0	1	35	9	0	54
% of the number of tweets on this day	16.7%				1.9%	64.8%	16.7%		
January 25, 2015	0	9	0	79	0	0	0	0	88
% of the number of tweets on this day		10.2%		89.8%					
January 28, 2015	0	0	0	0	0	40	0	0	40
% of the number of tweets on this day						100.0%			

Table 23. Distribution of the CDC tweets by type on peak days.

Date	Statement, announcement, report (% of the total number of tweets per period)	Press/news release, scientific publication, press conference, interview	Retweets	Tweeting from official events	Personal story, feature story	Responses and Twitter chats	Public health messages for the general public	Public health messages for health workers	Total
4 August, 2014	1	0	0	0	0	30	1	0	32
% of the number of tweets on this day	3.1%					93.8%	3.1%		
7 August, 2014	1	0	2	0	1	46	1	2	51
% of the number of tweets on this day	2.0%		3.9%		2.0%	90.2%	2.0%	3.9%	
8 August, 2014	2	0	0	0	1	36	1	1	41
% of the number of tweets on this day	4.9%				2.4%	87.8%	2.4%	2.4%	
2 October, 2014	1	0	0	0	0	49	2	0	52
% of the number of tweets on this day	1.9%					94.3%	3.9%		
8 October, 2014	3	0	3	0	0	30	1	0	37
% of the number of tweets on this day	8.1%					81.1%	2.7%		

Figure 61. Temporal distribution of retweets of the WHO and CDC Ebola tweets throughout the full study period.



5.8.2. WHO and CDC retweets

Figure 61 demonstrates the temporal distribution of the WHO and CDC Ebola retweets throughout the full study period. WHO tweets were retweeted 166,463 times by 71,932 unique Twitter users, whereas CDC tweets were retweeted 49,717 times by 27,976 users. As it can be observed in Figure 61, WHO tweets were mostly retweeted in August 2014, September 2014 and October 2014, whereas CDC tweets were highly retweeted in August 2014 and October 2014. The highest number of the WHO retweets was observed on 8 August 2014 (7,846 retweets) and 14 August 2014 (8,030 retweets), the period when WHO declared the Ebola outbreak a Public Health Emergency of International Concern. The highest number of CDC retweets was observed on 1 October 2014 (3,379), 2 October 2014 (3,531), the days following the first Ebola case diagnosis in the USA.

The top 10 most retweeted tweets posted by the WHO and their categorisation are presented in Table 24. Five of these tweets were “Public health messages for the general public”, four tweets were “Statements, announcements, or reports”, and one, the most retweeted tweet, belonged to the “Personal story, feature story” and was retweeted 3,195 times. Five of the original tweets were posted in the 3rd sub-period, four in the 4th sub-period, and one in the 5th sub-period.

Table 24. Top 10 most retweeted tweets posted by the @WHO account.

	Date	Sub-period	Tweet text	Number of retweets	Category
1	30 September 2014	4	He did not cry when his mother died of #Ebola because he didn't want to scare his brothers http://goo.gl/2TZaXR	3,195	Personal story, feature story
2	9 May 2015	5	WHO declares the end of the #Ebola outbreak in #Liberia Liberia #StaysVigilant	3,051	Statements, announcements, reports
3	14 August 2014	3	The highest #Ebola virus level is found in a dead body, hence currently the highest risk of Ebola transmission is during burial ceremony	2,388	Public health message for the general public
4	19 August 2014	3	Travel to and from #Ebola-affected countries is low-risk. Find out what you need to know in this infographic	1,996	Public health message for the general public
5	20 October 2014	4	WHO officially declares the #Ebola outbreak in #Nigeria over and commends the country on its diligence to end the transmission of the virus	1,905	Statements, announcements, reports
6	31 July 2014	3	There is no evidence base that #homeopathy can cure #Ebola. Severely ill patients require intensive supportive care http://goo.gl/E97CSQ	1,874	Public health message for the general public
7	8 August 2014	3	Bathing with salt and warm water, drinking water with salt does NOT cure #Ebola. Facts about what helps treat Ebola http://goo.gl/xeWzAU	1,848	Public health message for the general public
8	15 September 2014	3	#Ebola cases and deaths reported in #Nigeria and #Senegal as of 6 September 2014 http://t.co/kLrF6spESc	1,598	Statements, announcements, reports
9	20 October 2014	4	#Nigeria is now free of #Ebola virus transmission http://goo.gl/AMTL1s	1,454	Statements, announcements, reports
10	3 October 2014	4	Do NOT give aspirin or ibuprofen to #Ebola patients because they can make bleeding worse	1,340	Public health message for the general public

The top 10 most retweeted CDC tweets are presented in Table 25. Nine of them belong to the “Public health message for the general public “, including the most retweeted tweet, which was retweeted 4,155 times. One of the tweets belonged to the “Statements,

announcements, reports». Seven of the original tweets were posted in the 4th sub-period and three in the 3rd sub-period.

Table 25. Top 10 most retweeted tweets posted by the @CDCgov account.

	Date	Sub-period	Tweet text	Number of retweets	Type
1	30 September 2014	3	#Ebola is spread by direct contact w/bodily fluids of a sick person or exposure to contaminated objects, like needles	4,155	Public health message for the general public
2	12 October 2014	4	#Ebola is not spread through casual contact or through the air.	1,743	Public health message for the general public
3	23 October 2014	4	#Ebola is spread by direct contact w/ body fluids of a sick person or exposure to contaminated objects, like needles.	1,079	Public health message for the general public
4	1 August 2014	3	#Ebola poses no significant risk to US public. Only transmitted by contact with body fluids and only by someone w/ signs of illness.	868	Public health message for the general public
5	12 October 2014	4	#Ebola is spread by direct contact w/ body fluids of a sick person or exposure to contaminated objects, like needles.	846	Public health message for the general public
6	5 October 2014	4	Ebola is not spread through casual contact or through the air.	841	Public health message for the general public
7	17 October 2014	4	Get the facts about #Ebola. Here's what you need to know about when a person can spread the disease to others.	774	Public health message for the general public
8	16 October 2014	4	#Ebola virus is spread through direct contact with the blood or bodily fluids of a person who is sick with #Ebola.	738	Public health message for the general public
9	8 August 2014	3	#Ebola can only spread to others after symptoms begin. Get the facts:	699	Public health message for the general public
10	15 October 2014	4	2nd health worker exhibited no #Ebola signs/symptoms on Frontier Airlines flight 1143 on 10/13, but passengers should contact 1-800-CDC-INFO	680	Statements, announcements, reports

WHO and CDC “Public health messages for the general public” retweets.

Figure 62 demonstrates the number of WHO “Public health messages for the general public” retweets posted within the first three days of original tweet creation.

Figure 62. Number of the WHO “Public health messages for the general public” retweets posted within the first three days of original tweet creation.

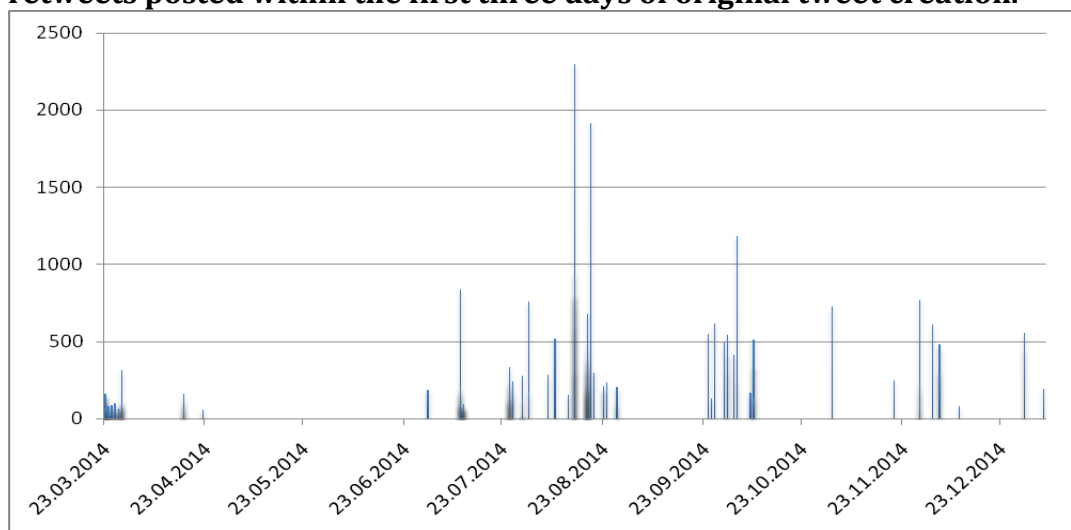


Table 26 demonstrates the most retweeted (more than 1,000 retweets within the first three days) WHO “Public health messages for the general public” tweets. 2 of them were posted in the 3rd and 1 in the 4th sub-periods. The first two were about EVD transmission, and the third was about EVD treatment.

Table 26. The most retweeted (more than 1,000 retweets within the first 3 days) WHO “Public health messages for the general public” tweets.

Date	Number of retweets (within 3 days of the original tweet creation)	Tweet text
14 August 2014	2,300	The highest #Ebola virus level is found in a dead body, hence currently the highest risk of Ebola transmission is during burial ceremony
19 August 2014	1,927	Travel to and from #Ebola-affected countries is low-risk. Find out what you need to know in this infographic http://t.co/xPW0BHH5gI
3 October 2014	1,183	Do NOT give aspirin or ibuprofen to #Ebola patients because they can make bleeding worse http://t.co/ESB4biRvpR

Figure 63 demonstrates the number of the CDC “Public health messages for the general public” tweets retweets posted within the first three days of original tweet creation.

Figure 63. Number of the CDC “Public health messages for the general public” retweets posted within the first three days of the original tweet creation.

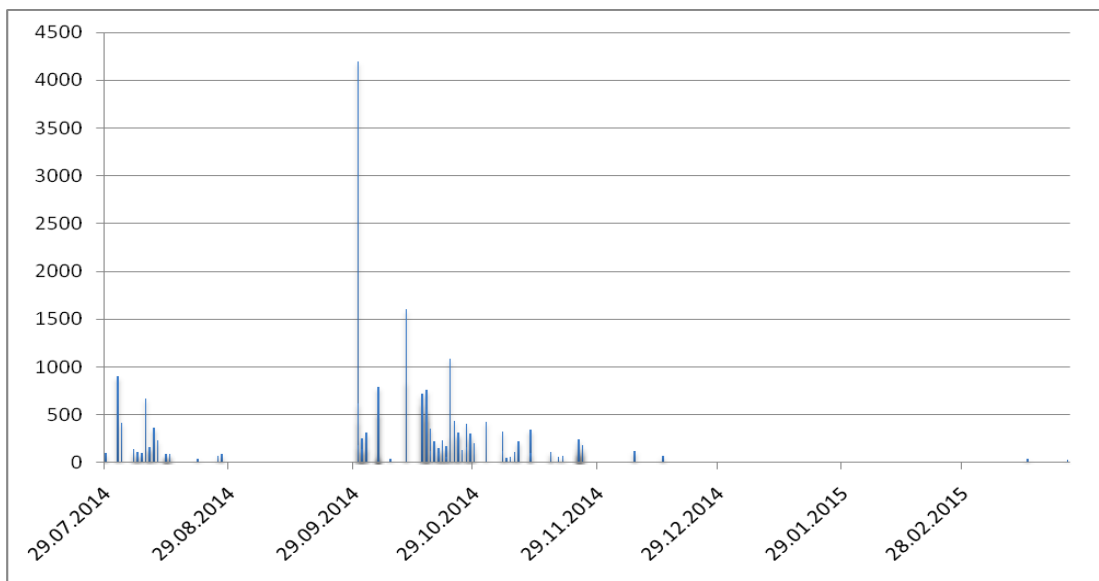


Table 27 demonstrates the most retweeted (more than 1,000 retweets within the first three days) CDC “Public health messages for the general public” tweets. All of them were posted in the 4th sub-period and contained the information that “Ebola is spread by direct contact with bodily fluids of a sick person or exposure to contaminated objects” and “is not spread through casual contacts or through the air”.

Table 27. The most retweeted (more than 1,000 retweets within the first 3 days) CDC “Public health messages for the general public” tweets.

Date	Number of retweets (within 3 days of the original tweet creation)	Tweet text
30 September 2014	4,195	#Ebola is spread by direct contact w/bodily fluids of a sick person or exposure to contaminated objects, like needles http://t.co/lVjLoWYBki
12 October 2014	1,601	#Ebola is not spread through casual contact or through the air. http://t.co/Ayq3gs9NVf
23 October 2014	1,083	#Ebola is spread by direct contact w/ body fluids of a sick person or exposure to contaminated objects, like needles. http://t.co/tsSjZ8g8nb

5.8.3. WHO and CDC mentions and replies

The WHO account was mentioned 202,852 times by 96,610 unique users, whereas 11,833 of these mentions were replies, which were posted by 5,772 users. The CDC account was mentioned 86,940 times by 48,869 users, 7,195 of these mentions were replies, posted by 4,796 users. Figure 64 demonstrates that the @WHO account was highly mentioned from August 2014 to November 2014, with the highest number of mentions (13,046) on 20 October 2014. It was the day when the WHO declared Nigeria free of Ebola virus transmission. The CDC account was highly mentioned in August 2014 and in October 2014, with the highest number of mentions on 30 September 2014 (7,003) and 16 October 2014 (6,859). 30 September 2014 was the day when the first Ebola case was diagnosed in the USA. On 16 October 2014 the news that the nurse who cared about Ebola patient in the USA was allowed on a commercial flight having a slight fever shortly before the diagnosis of the EVD was discussed on Twitter mentioning the CDC account.

Figure 64. Temporal frequency distribution of @WHO and @CDCgov mentions throughout the full study period.

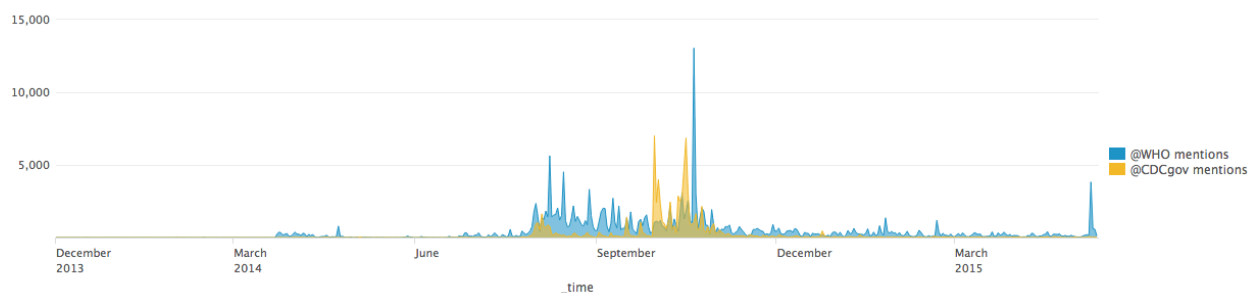
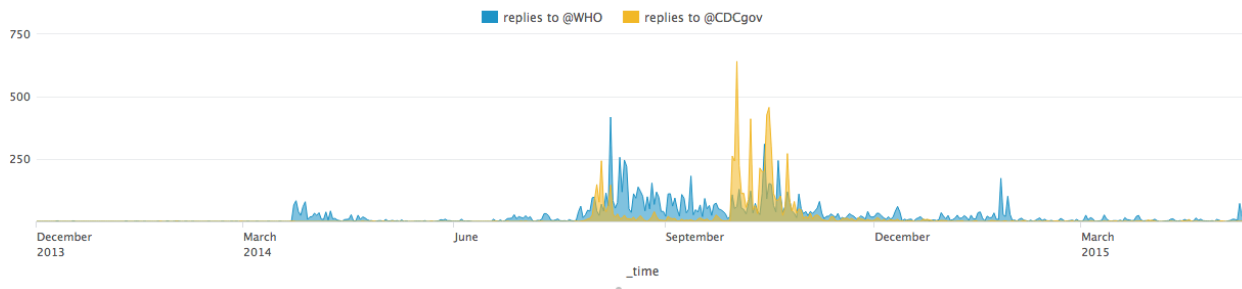


Figure 65 demonstrates that among the mentions, the highest number of replies was received by the WHO on 8 August 2104 (418 replies), the day when Ebola was declared a Public Health Emergency of International Concern. The highest number of replies to CDC was posted on 2 October 2014 (641 replies), the day when CDC held one of their Twitter chats.

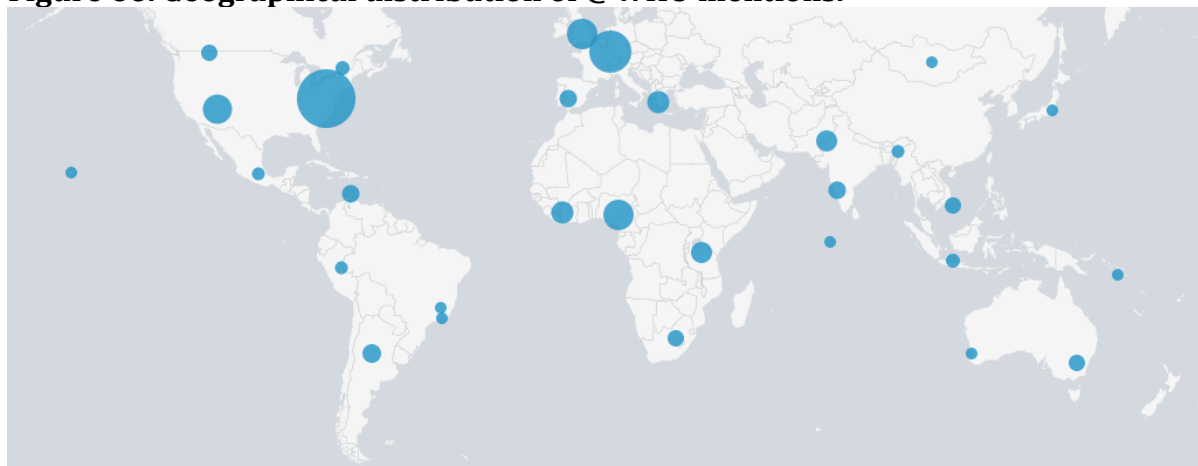
Figure 65. Temporal frequency distribution of replies to the @WHO and @CDCgov throughout the full study period.



Geographical distribution of the @WHO and @CDCgov mentions.

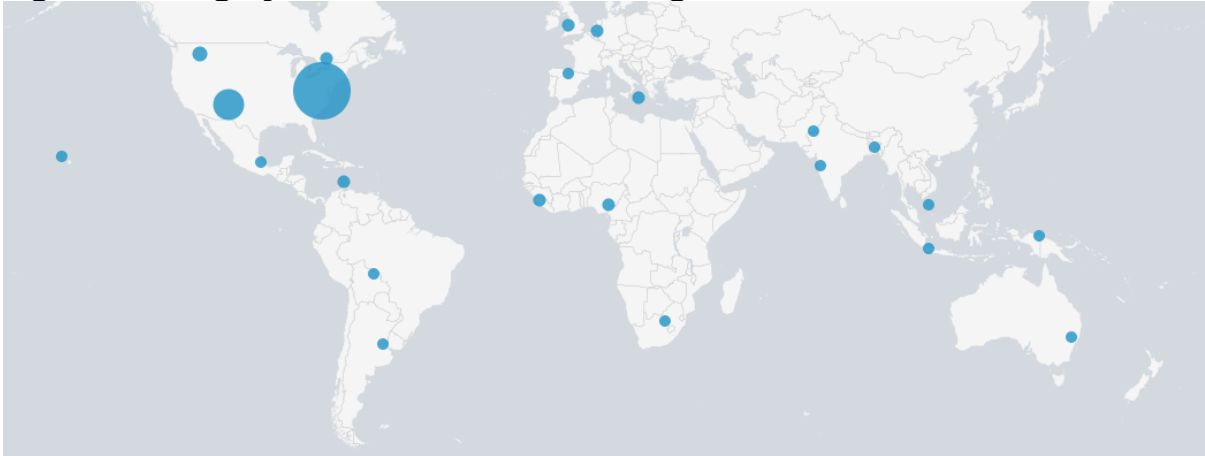
2,253 (3.1%) of the @WHO mentions were geo-tagged. Figure 66 demonstrates that the @WHO mentions came from all over the world, with the highest number of message coming from North America, Europe, and West Africa.

Figure 66. Geographical distribution of @WHO mentions.



1,395 (1.6%) of @CDCgov mentions were geo-tagged. Figure 67 demonstrates, that @CDCgov mentions came mostly from North America.

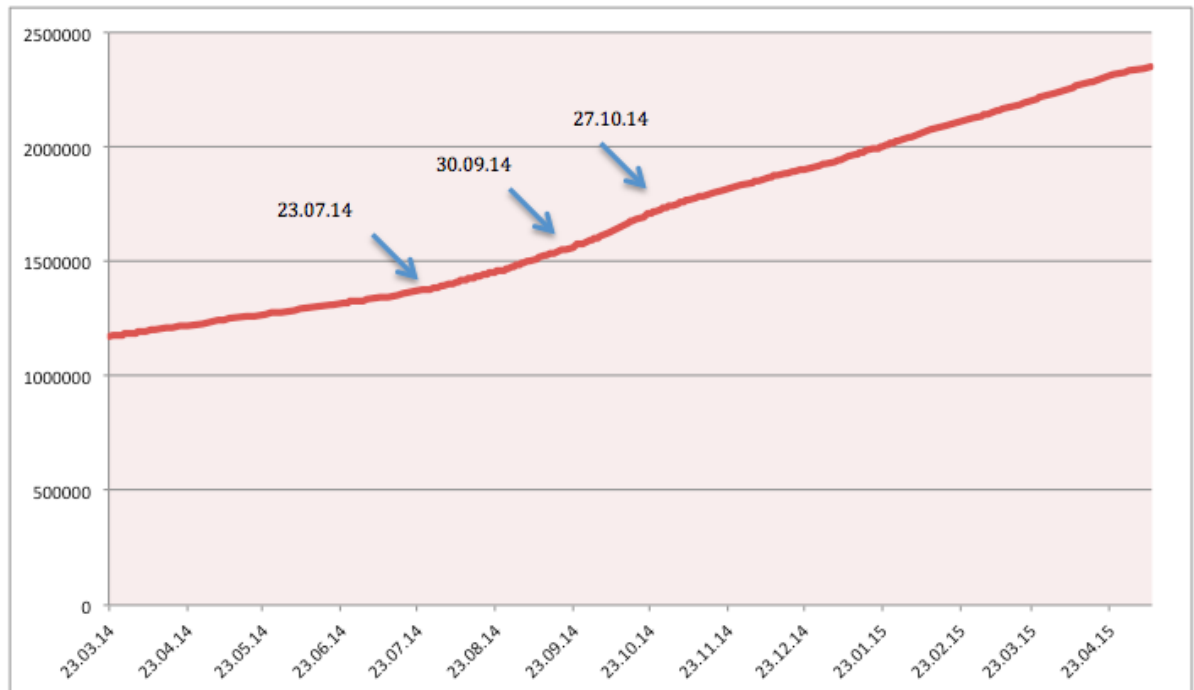
Figure 67. Geographical distribution of @CDCgov mentions.



5.8.4. Followers of the WHO and CDC accounts

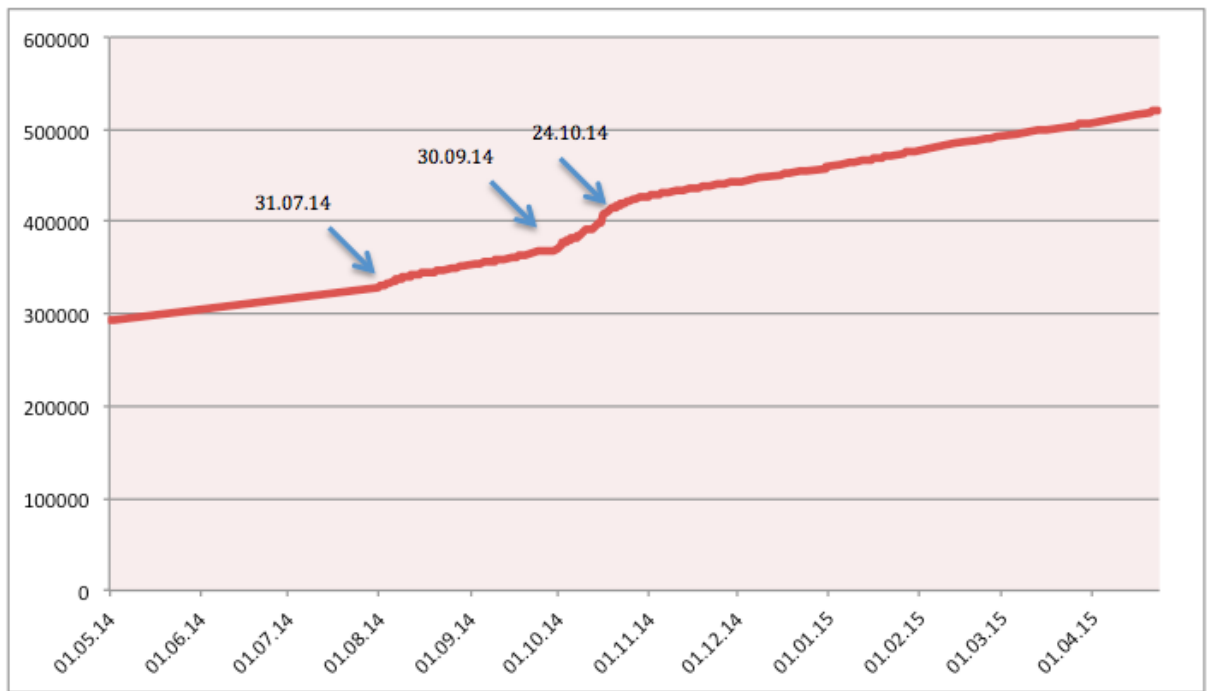
At the beginning of the study period the number of the @WHO account followers was 1,173,156, whereas it had reached 2,351,184 by the end of the study period (Table 19). Figure 68 demonstrates the growth of the WHO account follower number over the study period. It can be seen, that the growth of the follower number increased on 23 July 2014 (the beginning of the 3rd sub-period) and 30 September 2014 (the beginning of the 4th sub-period), and decreased again on 27 October 2014 (the end of the 4th sub-period). The correspondence of the @WHO follower number growth level with the study sub-periods (and therefore with Ebola news events) might suggest that it was changing due to the changing interest level of the Twitter users in Ebola topic.

Figure 68. Growth in the number of followers of the @WHO account over time.



At the beginning of the study period the number of the @CDCgov followers was 292,595, whereas it had reached 520,508 by the end of the study period (Table 19). Figure 69 demonstrates that the @CDCgov follower number growth increased at the end of July 2014 (the beginning of the 3rd sub-period) and dramatically increased on 30 September 2014 (the beginning of the 4th sub-period), and decreased again on 24 October 2014 (the end of the 4th sub-period). However, the number of records before 31 July 2014 was low (just 4 tweets), which may cause the first visual increase in the growth level. The correspondence with the study sub-periods (and therefore with Ebola news events) might suggest that it was changing due to the changing interest level of the Twitter users in Ebola topic.

Figure 69. Growth in the number of followers of the @CDCgov account over time.



Chapter 6. Discussion and Conclusion

In analysing more than 60 million tweets posted by more than 12 million unique Twitter user accounts within the period from 1 December 2013 to 12 May 2015, I was able to discern the temporal dynamic and engagement of Tweets in relation to the West African Ebola outbreak. I also revealed the ways in which two key governmental health organisations used Twitter to disseminate information and engage the public.

6.1. Pattern of Twitter use in the general Twitter population

To answer the RQ1 about the pattern of Twitter use in the general Twitter population during the West African Ebola outbreak, I examined the overall features of the dataset such as language of tweets and their geography, as well as the temporal tweet frequency distribution over the studied period and its correspondence with the EVD epidemiology and Ebola news events.

Languages of tweets

My analysis has revealed that tweets in the dataset were posted in 149 languages. Tweet filtering criteria were limited to Latin script, which means that languages such as Russian, Arabic, and Japanese, were not included in the dataset. However, as has been demonstrated previously, in many countries people tweet in English in order to reach wider audience even if this is not the official country language. (28) Therefore, some of the tweets from the countries that use a different alphabet could have been included in the study dataset. In several languages with a non-Latin alphabet, such as Indonesian and Malay, the word Ebola is still spelled as “Ebola”. Therefore, tweets from these countries were also included in the study. Interestingly, the number of identified languages in the

dataset was still relatively high. For example, in a study of non-thematic tweets 78 languages were identified covering the period between 2010 and 2012. (28) This can be explained by the difference of the methods we used to identify them. In the latter study, authors used the languages reported by Twitter users, whereas I used the language augmentation service of DataSift that examines the tweet text itself. In the study dataset the most prevalent languages were: English, Spanish, Portuguese, French, and Indonesian, whereas, in the study of non-thematic tweets the most prevalent languages were English, Spanish, Indonesian, Malay, and Japanese. (28) Although Japanese tweets were not included in the study dataset, Indonesian and Malay were. The higher prevalence of English and European languages suggests that the Ebola Twitter conversation was not as prevalent in Asian countries as in English speaking, European, and, possibly, Latin American countries.

Overall geography of tweets

1.2% of the tweets in the study dataset were geo-tagged, which is similar to the prevalence of geo-tagged tweets in the other large datasets of non-thematic tweets that had 1.2% (25) and 1.6% (36) of them. Geographical mapping of the geo-tagged tweets demonstrated that most of them were posted from North America and Europe. This matches the results of the tweet language distribution in my study. Comparing this with the fact that, globally, tweets are mostly posted from North America, Europe, Latin America, Asia, and the Middle East, (25,36) and considering that the tweets in my thematic dataset contained only Latin script of the keyword and hashtags (therefore excluding tweets from most Middle Eastern countries), the results of this study demonstrate that the global Ebola conversation on Twitter was less prevalent in Asian and in Latin American countries than in the general Twitter conversation, but was rather shifted to North America and Europe.

Twitter activity, EVD epidemiology, and Ebola news events

The Ebola conversation on Twitter did not correspond with EVD epidemiology (i.e. incidence, case fatality rate, or infection rate). However, and not surprisingly, tweets did correspond with Ebola news events. The Twitter conversation was initiated on 22 March 2014, a day when a mysterious haemorrhagic fever in Guinea was confirmed as Ebola, which was followed by the release of the first WHO report concerning the West African Ebola outbreak. In this period following this event a relatively high prevalence of tweets was posted from West African countries, especially from Guinea, Liberia, and Sierra Leone, as well as from North America and Europe. The frequency of tweets increased on 23 July 2014, two days before the first EVD case was tested and officially confirmed in Lagos, Nigeria. The tweet frequency was rapidly increasing further when two US missionary workers infected with Ebola were evacuated to the USA from Liberia at the beginning of August 2014 and when the WHO declared the Ebola outbreak a Public Health Emergency of International Concern. In this period a relatively high prevalence of tweets was also posted from West Africa, especially Nigeria, as well as from North America and Europe. Tweet frequency reached its peak at the beginning of October 2014, after the first EVD case was diagnosed in the USA on 30 September 2014 and demonstrated significant peaks throughout October 2014, when the first Ebola infected patient died (8 October 2014) and when three more EVD cases were diagnosed in the USA (Table 1). The vast majority of tweets in this period were posted from North America. Although existing evidence on the Ebola conversation on Twitter is scarce, the findings of my study match previous findings that have demonstrated that it did not correspond with the EVD epidemiology, but was initiated and amplified by the Western media coverage. (5,97,99–102,167) The news of the first Ebola cases diagnosed in Nigeria and transported and diagnosed in the USA had the highest influence on the temporal tweet frequency distribution, which was disproportionate to the actual scale of the outbreak in these countries (20 reported cases in

Nigeria and 4 reported cases in the USA, compared to thousands of cases in Guinea, Liberia, and Sierra Leone). (89) However, the news of Ebola cases diagnosed and evacuated to other European countries such as Spain, and the UK did not cause high peaks in Twitter Ebola conversation, most likely due to a lower number of Twitter users in these countries or to a lower global media coverage. Interestingly, the Twitter conversation started to gradually drop since the beginning of November 2014, after all four cases had been diagnosed in the USA. This suggests that the public interest in the topic dropped off with time, perhaps due to less intense Ebola coverage by the Western media. Even when an Ebola infected patient was transported to the USA from Sierra Leone on 13 March 2015, just a slight tweet frequency increase could be observed on the temporal tweet frequency distribution. As reported by Lee, (79) who compared the sentiments of tweets on the days when the first Ebola patient was transported to the USA (1 August 2014) and 13 March 2015, the prevalence of negative tweets decreased by 14.8%. He suggests that the public “fear-factor” decreased, while the “indifference-factor” increased. I also demonstrated that since 1 November 2014 most tweets were posted from North America and Europe, starting to match the overall trend of the dataset. This also supports the argument that public anxiety in this period dropped and the interest of the general Twitter public in Ebola conversation decreased.

I have demonstrated that in the period from 22 March 2014 and 30 September 2014 a relatively high prevalence of tweets was posted from West Africa. Although the number of social media users in the countries of this region continues to grow, (94,95) it is not clear whether tweets from this area were posted mainly by local citizens or by international workers who were assisting with the Ebola outbreak. It is also possible that a significant proportion of tweets from this location were posted by spam accounts (such as automated tweet aggregators). Therefore, it is difficult to tell whether the messages on the platform were actually posted by the most affected populations and, conversely, reached them

through the platform. More research is needed in order to assess the content of the tweets from West Africa (region of Guinea, Liberia, and Sierra Leone) and examine who was posting them. Moreover, the absolute number of tweets from West African countries was relatively high even in October 2014, but their prevalence was overtaken by the vast number of tweets posted from North America. This can be explained by the continuous and extensive coverage of Ebola throughout October 2014, by the fear that the US population expressed when four Ebola cases were diagnosed in their country, (5) and by the fact that Twitter is most prone to express the reaction of the American public since 23% of all registered Twitter accounts come from the USA. (168) It also demonstrates that the world's attention was brought to the Ebola epidemic only once the first cases were diagnosed outside of West Africa at the end of September 2014.

6.2. Twitter engagement

To address the RQ2 about general Twitter engagement, I calculated the prevalence of mentions, retweets, and replies in the study dataset to understand how Twitter users were engaged in the Twitter Ebola conversation. To further identify accounts that demonstrated the highest level of Twitter engagement related to the West African Ebola outbreak and the place of governmental and non-governmental health organisations among them, I examined the top 20 most retweeted and mentioned⁴⁸ accounts, the top 20 accounts which received the highest number of replies, and the top 10 retweeted tweets and their authors over the whole study period and within each of the 5 sub-periods.

General Twitter engagement

In my examination of how Twitter users were engaged in the Twitter Ebola conversation, I found that the overall prevalence of any mentions⁴⁹ in my dataset was 61%.

⁴⁸ Mentions without retweets

⁴⁹ Any mentions include retweets, replies, and all other mentions

Whereas, studies of non-thematic tweet datasets reported the prevalence of mentions at 50% (25) and 55.1%. (36) In my dataset 25.4% of unique users were mentioned over the whole study period, whereas in a study on a non-thematic dataset 12.4% of users were mentioned. (36) The higher prevalence of any mentions (including retweets) and mentioned users in my datasets compared to non-thematic dataset suggests that Twitter users were overall more engaged in the Ebola conversation and that they were more likely to share information with other Twitter users in comparison with the general Twitter conversation.

48.3% of the tweets in my dataset were retweets. This is almost twice as high as that in the large datasets of non-thematic tweets, which reported the prevalence of retweets at 27% (25) and 23.9%. (36) The overall proportion of retweeted users among unique tweeting users in my study was 12.8%, whereas the study of non-thematic tweets reported that 4.5% of unique users in 2012 were retweeted. (36) Moreover, the prevalence of retweets increased after the Ebola conversation started and reached its peak in October 2014 (55.8%), after which it dropped to 42.8%, which is still relatively high in comparison with the period before the conversation was initiated (32.3%). A similar pattern was observed with the prevalence of retweeted and retweeting users among all the unique users in the dataset, which increased after the conversation started and reached their peaks in October 2014, with retweeting users at 74.5% and retweeted users at 12.8%. In every studied period more than half of retweets contained URLs, with the exception of October 2014. Although the prevalence of retweets was the highest then, the prevalence of retweets containing URLs among them was only 19%. This is still higher than that observed in non-thematic datasets at 12% (25) and 15.9%. (36) This suggests that Twitter users were more likely to retweet information and news during the Ebola outbreak in comparison with the general Twitter conversation. In October 2014, however, although

users were frequently retweeting other users, their tweets were less likely to contain the links to news stories or other information articles.

The prevalence of mentions excluding retweets in my dataset was 12.8% and the prevalence of tweets containing replies was 6.6%, whereas in a non-thematic dataset in July 2014 the prevalence of replies was 24%. (25) The prevalence of tweets containing replies and mentions without retweets in the dataset also decreased in comparison with the period before the Twitter Ebola conversation started. Considering the fact that mentions without retweets and replies demonstrate direct engagement with other Twitter users, this finding suggests that when the retweeting level was relatively high, direct engagement between Twitter users with regards to Ebola topic was relatively low. These findings also indicate that the engagement parameters differ in thematic datasets in comparison with non-thematic datasets due to a variety of thematic and contextual factors, as it has previously been shown. (136)

Accounts demonstrating highest Twitter engagement level

In order to discern the key players involved in the Twitter discussion surrounding the West African Ebola outbreak, I identified accounts that achieved the highest level of engagement with other Twitter users during this period. I demonstrated that nine among the top 20 retweeted accounts, 11 of the top 20 mentioned⁵⁰ accounts, and seven among the top 20 accounts that received the highest number of replies belonged to news organisations. This suggests that people were sharing news stories and actively sought information on Twitter, and supports the fact that Ebola conversation on Twitter corresponded with news events. This is consistent with the findings of the study that investigated tweets about 2009 H1N1⁵¹ pandemic, (169) which demonstrated that more than half of the tweets contained news and information material. Moreover, during that

⁵⁰ Mentions in the top 20 mentioned accounts exclude retweets

⁵¹ Swine flu

outbreak reputable news organisations were more successful in risk communications, providing accurate information, than public health agencies themselves. (65) This highlights the importance of Twitter as a tool for rapid communication and information sharing during epidemics. This is also a lesson for health authorities, that in future outbreak scenarios they could consider tracking the information provided by these organisations and collaboration with them.

Interestingly, six of the top 20 retweeted accounts were humour accounts. Their number was especially high in the study period over October 2014 (ten of the top 20 retweeted accounts), which might explain why the prevalence of retweets containing URLs in this period was relatively low. Users were sharing humorous content rather than news and information. Six of the top 20 accounts with the highest number of replies were also humour accounts. In October 2014 humour accounts, together with accounts of news organisations, had the leading position among the top 20 accounts that received the highest number of replies. This suggests that Twitter users were directly engaging with humour accounts to a high degree in October 2014. This observation also corresponds with the findings from a 2009 H1N1 pandemic tweets study, which also reported high prevalence of humorous tweets. (169) However, the latter study also demonstrated that their prevalence dropped in the periods when major news events happened and the perceived seriousness of the situation increased. This was not the case in my study: humour accounts were especially highly retweeted in October 2014 when four Ebola cases were diagnosed in the USA, the country where most of the tweets of that period came from. Although there are a number of humour accounts, which have millions of followers on Twitter, (170) the reasons for their popularity on social media have not been studied much. Their high popularity might be explained by the fact that among the main purposes social media is used for are passing time, entertainment, and relaxation. (12) Several big brands use humour in their brands' social presence in order to attract attention, become

memorable, develop the brands' personality, and connect with social media users on an emotional and human level. (171) Social media is a medium where it is possible to get away from a routine daily life. "It is an escape from reality", and it takes your mind off things. (172) Perhaps this might explain the popularity of humour content and accounts during emergencies, when the public experiences fear and anxiety. Although some of the humour accounts in this study had already been popular on Twitter, having had high number of followers before they started to participate in the Ebola conversation, which might explain why they were highly retweeted, some of the most retweeted humour accounts were created during the study period, with Ebola as a specific focus of their tweets, and did not have many followers when their first tweets were posted. The appearance of humour and especially bad jokes following any catastrophe has been reported, (173) but has not been well studied. Humour accounts' popularity in this study might be explained by the general popularity of Ebola as a topic for Twitter conversation, and by the increased concern of the general public and Twitter users on this issue. Reading and sharing humour content on Twitter might have acted as a means for relief and tackling anxiety, as it is known that humour, even dark humour, is used in stressful situations as a buffer. (174,175) It has also been suggested that being able to laugh at traumatic events does not cause ignorance of these events, but enables resilience in the face of a potential threat. (176) Nevertheless, I have not examined what the actual content of tweets posted by humour accounts was; it might have not always been humour content. For example, as has been previously reported, one of these accounts retweeted a public health information tweet posted by CDC, which received a higher number of retweets than the most retweeted tweet posted by CDC itself. (8) This can be explained by a higher number of followers that this humour account had in comparison with the CDC account. Therefore, the issue of content posted by humour accounts should be investigated further, and their ability to spread public health or other accurate information should be further explored.

A high number of Twitter celebrities were among the top 20 most mentioned accounts throughout the study period, and five of the top 10 retweeted tweets were posted by them, suggesting that these accounts have high influence in the Twittersphere during the outbreak periods. I have not investigated who the owners of these accounts were, but I identified that they included health specialists, journalists, or bloggers, who actively use Twitter for communication with the public. Some of the most retweeted tweets posted by these accounts contained humorous content and personal opinions. The potential of Twitter celebrities to spread public health information could be further explored.

Three of the top 10 most retweeted tweets in the dataset were published by celebrities, most likely due the high number of followers and reputation they have on Twittersphere, which also demonstrates their potential to spread the information during public health emergencies. Many UN organisations already collaborate with celebrities, appointing them as goodwill ambassadors in order to attract attention and focus the world's eyes on organisations' goals. (177–179) However, the collaboration of health authorities with celebrities in order to share important information on social media during emergencies should be further explored as well as the content they share and the impact it might achieve. For example, in the study dataset some of the most retweeted tweets of celebrities contained a call to donate money to fight Ebola. Similarly, during the 2015 Nepal earthquake, Bollywood celebrities used their Twitter accounts to share emergency contact numbers in Nepal. (180) But the impact of these messages, such as how many people donated the money or used these contacts, has not been measured.

Public health organisations may consider the possibilities of closer collaboration with the accounts of news organisations, Twitter celebrities, celebrities, and humour accounts that have such a high reach and level of engagement with the general Twitter population during public health emergencies, especially in the periods when the public expresses high levels of anxiety in the case of humour accounts. They also should consider strategic use of

humour and emotional? content, where feasible, in their own tweets for a closer engagement with their Twitter audience and for a wider dissemination of important information.

Health authorities

Twitter accounts of several governmental and non-governmental health organisations also achieved a high level of Twitter engagement throughout the epidemic. Accounts of the WHO and UNICEF were among the top 20 most retweeted accounts, accounts of the WHO and CDC were among the top 20 most mentioned accounts, and accounts of the WHO, CDC and the White House were among the top 20 accounts that received the highest number of replies. This suggests that Twitter users turned to these accounts in order to seek the information during the West African Ebola outbreak. Whereas, the WHO and UNICEF accounts were the sources of information for the global community, the accounts of CDC and White House provided the information for the US public, which highly influenced the Twitter Ebola conversation when Ebola cases were transported or diagnosed in the USA.

However, none of the top 10 most retweeted tweets was posted by these organisations which indicates that none of their tweets managed to receive the highest levels of attention. Five of the top 10 most retweeted tweets in the dataset were posted by Twitter celebrities, three by celebrities, one by an individual, and one by a humour account. Although the popularity of celebrity and humour tweets in this case might be explained by high numbers of followers these accounts had since the beginning of the outbreak (more than 20 million for celebrities and nearly 767,000 followers for a humour account), whereas the account of the WHO had 1.2 million followers and the CDC account had nearly 293 thousand followers, it was not the case for other six most retweeted tweets. The number of followers of Twitter celebrities and an individual account that posted these tweets ranged from 500 to 45,000 at the beginning of the outbreak. This suggests, that

their tweets were popular due to their content, rather than the initial account popularity. Some of these tweets contained humour, others expressed personal opinions. Nevertheless, in the period after the WHO published its first report on Ebola and before EVD spread to Nigeria, two of the most retweeted tweets were published by the WHO and the UN with the reference to the WHO. They contained public health messages. This suggests that these organisations do have a potential to be widely heard on Twitter and that adoption of new strategies such as wise sharing of humorous or emotional content may increase the reach of their tweets.

It has been reported that the WHO and CDC used Twitter in addition to their traditional communication channels in order to provide accurate information about Ebola. (4) These two organisations were among the top 20 mentioned overall in my study. This corresponds with the finding of another study where these accounts were among the top 10 mentioned accounts on Twitter among the messages about Ebola in September 2014. (7) Whereas the WHO account was highly retweeted, mentioned and received a high number of replies throughout the study period, the CDC account was highly mentioned and received high number of replies mainly between 23 July 2014 and 1 November 2014, in the period when Ebola cases were firstly transported to the USA and when they were further diagnosed in the USA. This suggests that, not surprisingly, the account of the WHO, a health organisation that operates globally, achieved a high level of engagement unrelated to the geography of Ebola news, whereas the account of CDC, the US organisation, demonstrated the highest level of engagement when major Ebola news came from the USA. Twitter users globally referred to the WHO Twitter account when they sought information about Ebola, whereas the CDC Twitter account was mainly heard by the US public.

6.3. WHO and CDC Twitter Ebola activity

To answer the RQ3 about the ways the WHO and CDC used Twitter during the West African Ebola outbreak and the level of engagement they achieved with the general Twitter population, I performed a content analysis of all the tweets posted by these organisations during the study period, as well as identified the periods of their highest Twitter activity. I further looked at the engagement measurements, such as retweets, mentions, replies, and the number of their followers throughout the study period.

Ebola related Twitter activity of the WHO and CDC

I found that the highest prevalence of tweets posted by these organisations (29.9% of the WHO tweets and 32.8% of the CDC tweets) comprised direct responses to other Twitter users, as a part of Twitter chats or independently, suggesting that these organisations proactively responded to Ebola on Twitter and put an emphasis on direct engagement with their Twitter audience, listening to their concerns and replying to their questions during the outbreak. Both the WHO and CDC posted Ebola tweets containing official statements, announcements, reports, news releases, tweets from press conferences, and tweets containing personal or feature stories. A significant number of tweets contained public health messages for the general public and health workers. The WHO also posted tweets from official events on Ebola, which their staff participated in. A significant proportion of CDC tweets were retweets of accounts of other organisations or CDC officials. This demonstrated a variety of strategies that already have been employed by these organisations for the communication during the outbreaks.

To understand more clearly how these organisations used Twitter during an outbreak scenario, I identified the days when they posted the highest number of tweets and examined the topics of the discussion. The highest level of the WHO Twitter activity was observed when the organisation communicated the information about a rapidly evolving

Ebola outbreak in Guinea (24 March 2014), when the WHO declared the Ebola outbreak a Public Health Emergency of International Concern (8 August 2014), when the organisation was covering meetings in which the WHO staff took part at live on Twitter, and when the WHO organised its Twitter chat. In contrast, the highest level of the CDC Twitter activity was observed during its four Twitter chats on Ebola. Two of them were organised when the first patients infected with Ebola were evacuated to the USA, two other were organised after the first Ebola case was diagnosed in the USA. This demonstrates that the organisations put priority on communicating information in different periods. WHO, being a global organisation, announced major global events related to the outbreak, whereas CDC, which is primarily focused on the USA, actively communicated information when EVD was in the USA and when the US public was concerned about it.

By categorising the content of their tweets, I was able to discern that the overall prevalence of the tweets containing public health messages posted by the WHO was 8.6%, with the highest prevalence of them in a period after the organisation posted its first statement on Ebola (22 March 2014 – 21 July 2014). The overall prevalence of the tweets containing public health messages posted by the CDC was 12.8%, with the highest prevalence of them in a period when four Ebola cases were diagnosed in the USA and after that (30 September 2014 – 12 May 2015). This suggests that the WHO put an emphasis on communication of public health information to the global public when EVD was identified as the cause of the epidemic emerging in West Africa, whereas CDC started to communicate the public health information when Ebola-infected patients were evacuated to the USA and increased doing so when Ebola cases started to be diagnosed in the USA.

Twitter engagement of the WHO and CDC

I further examined the engagement of the general Twitter population with the WHO and CDC on Twitter, identifying the Ebola news events that made the general Twitter public turn to the WHO and CDC Twitter accounts, which are summarised in Table 28.

Table 28. Ebola news events that made the general Twitter public turn to the WHO and CDC Twitter accounts.

	WHO	CDC
Highest number of retweets	<ul style="list-style-type: none"> • WHO declared the Ebola outbreak a Public Health Emergency of International Concern 	<ul style="list-style-type: none"> • First Ebola case diagnosed in the USA
Most retweeted tweet	<ul style="list-style-type: none"> • First Ebola case diagnosed in the USA 	<ul style="list-style-type: none"> • First Ebola case diagnosed in the USA
Other tweets within top 10 most retweeted tweets	<ul style="list-style-type: none"> • First Ebola case was tested in Nigeria • First patients infected with Ebola were evacuated to the USA • WHO declared the Ebola outbreak a Public Health Emergency of International Concern 	<ul style="list-style-type: none"> • The period when four Ebola cases were diagnosed in the USA
Most retweeted tweets containing public health messages	<ul style="list-style-type: none"> • WHO declared the Ebola outbreak a Public Health Emergency of International Concern • First Ebola case diagnosed in the USA 	<ul style="list-style-type: none"> • The period when four Ebola cases were diagnosed in the USA
Highest number of mentions	<ul style="list-style-type: none"> • WHO declared Nigeria Ebola free 	<ul style="list-style-type: none"> • First Ebola case diagnosed in the USA • A second Ebola-infected health worker in Dallas took a flight a day before she was diagnosed with Ebola
Highest number of replies	<ul style="list-style-type: none"> • WHO declared the Ebola outbreak a Public Health Emergency of International Concern 	<ul style="list-style-type: none"> • Third CDC Twitter chat about the first Ebola case diagnosed in the USA
Significant increase in the number of followers	<ul style="list-style-type: none"> • First Ebola case was tested in Nigeria • First patients infected with Ebola were evacuated to the USA • First Ebola case diagnosed in the USA 	<ul style="list-style-type: none"> • First patients infected with Ebola were evacuated to the USA • First Ebola case diagnosed in the USA

The public turned to the WHO Twitter account when the first Ebola case was tested in Nigeria, when the first two Ebola infected patients were evacuated to the USA, when the WHO declared the Ebola outbreak a Public Health Emergency of International Concern, when the first Ebola case was tested in the USA, and when the WHO declared Nigeria Ebola free. In contrast, as was mentioned earlier, the highest level of the WHO Ebola related Twitter activity was observed in the period shortly after it was confirmed that the outbreak

in Guinea was caused by EVD, when the WHO declared the Ebola outbreak a Public Health Emergency of International Concern, and when it organised a Twitter chat with an Ebola survivor at the end of January 2015. This demonstrates disproportionality in the time frame of the information provision by the WHO and of the general Twitter public seeking the information. For example, in October 2014, when Ebola cases were tested in the USA and the US public expressed fear and anxiety, WHO did not demonstrate a high level of Twitter activity related to Ebola. Nevertheless, in this period the public also turned to the CDC Twitter account and CDC quickly responded to them. Thus, the highest level of the CDC Twitter activity was related to its four Twitter chats on Ebola, which were organised at the beginning of August 2014, when Ebola-infected patients were transported to the USA, and in October 2014, when four Ebola cases were diagnosed in the USA. The highest prevalence of the tweets containing public health messages was also observed at this time.

The most retweeted WHO tweet was posted on the day when the first EVD case was diagnosed in the USA. It contained an emotional personal story and was retweeted 3,195 times by the end of the study period. The fact that a personal story generated more Twitter engagement than public health messages conveys that the Twitter public demonstrates higher engagement with emotional and humane content, rather than messages containing official information. Nevertheless, five of the top 10 most retweeted WHO tweets contained public health messages for the general public and were posted in the period when the first Ebola case was diagnosed in Nigeria, when two first infected patients were transported to the USA, and when the WHO declared the Ebola outbreak a Public Health Emergency of International Concern. These were also the periods when the Twitter public demonstrated high engagement with the WHO account overall, and was seeking accurate public health information on Ebola. Moreover, nine of the top 10 most retweeted CDC tweets contained public health messages, as well as the most retweeted tweet that was also posted on 30 September 2014 and was retweeted 4,155 times. Seven of the most retweeted CDC tweets

were posted in the period when four Ebola cases were diagnosed in the USA, which also suggest that public was seeking accurate public health information from the organisation in that period. Interestingly, the most retweeted tweet in the whole dataset was posted by a Twitter celebrity, who had just around 2,500 followers. This tweet contained a political statement expressed as a joke and was retweeted 62,739 times (Table 14). Considering the relatively low number of followers that the account had, this tweet was likely retweeted due to its controversial content that attracted people's attention. This suggests that the combination of emotional content with important public health messages could help public health organisations to achieve even higher levels of engagement with the Twitter public. Nevertheless, more research is needed in order to investigate what kind of content might lead to the wider dissemination of tweets in public health emergencies.

Although it was not possible to study the geography of the WHO and CDC retweets in order to see how far they were spread, due to the absence of geo-tags within retweet metadata, it was possible to map the WHO and CDC mentions. I assumed that these results can demonstrate the geography of the audience of the WHO and CDC Twitter accounts. I have shown that the WHO mentions came from all over the world, and that their amount was especially prevalent in North America, Europe, and West Africa. The countries in these regions were highly affected or involved in the response to the West African Ebola outbreak. This suggests that the citizens of these regions were mostly aware of the outbreak, and that they turned to the WHO Twitter account for the information. CDC mentions came mostly from North America, demonstrating that US citizens turned to the CDC account to seek information about Ebola.

6.4. Methodological contribution

In addition to addressing my RQs, this research also demonstrated a way of studying thematic Twitter datasets that correspond with news events. Firstly, I have demonstrated a

way of studying a dataset covering a long period of time. Secondly, I demonstrated the possibility of using URLs in order to identify main topics of discussion in the days of interest.

Studying a dataset covering a long period of time

I have demonstrated a way of studying a dataset covering a long period of time by breaking it down into smaller sub-periods based on the tweet frequency distribution. It was relatively easy to identify the borderline dates between periods when the level of the conversation increased, and the events responsible, as in this dataset the increases happened abruptly. However, it was more difficult to choose the borderline date between the periods when the conversation started to gradually decrease. It is more challenging to say when exactly the conversation started to drop, and why, as it might have happened due simply to other world events that attracted users' attention, or due to the decreasing popularity of an issue among the public (fatigue), or for other reasons that it may not be possible to identify. In the case with Ebola, the Twitter conversation probably started to decrease when the Western media stopped their disproportionate coverage of the outbreak after the fourth Ebola case was diagnosed in the USA on 23 October 2014, which caused a decrease of the "fear-factor" and an increase of the "indifference-factor". This method enabled me to analyse each of the sub-periods separately examining all their features independently as analysing them as a whole would have left some of the observations neglected.

Use of URLs

A further methodological contribution relates to my use of URLs. 54.5% of the tweets in the dataset contained URLs which is higher than it was reported in non-thematic tweet datasets, at 15.9% in 2012 (36) and at 12.0% in 2014. (25) It has already been

demonstrated that it is possible to use URLs in order to identify the geographical focus of a tweet content. (36) I demonstrated that it is possible to use URLs to identify the main topic of discussion within a sample of tweets in peak days (days with the high level of tweet frequency) when more than half of the tweets contained URLs in them and more than half of the unique users shared the URLs.

Nevertheless, both of these approaches are based on the assumption that media coverage was a driver of the Twitter conversation. The long study period was divided based on the tweet frequency distribution, which I assumed to be dependent on media coverage. The peaks of interest in the temporal tweet frequency distribution I also assumed appeared due to particular news stories. Moreover, I did not identify and eliminate spam tweets from the dataset, which may have distorted the results of this method.

6.3. Strengths and limitations

To my knowledge, this is the first study to investigate Ebola conversation on Twitter over a long period of time (528 days) from the beginning of the outbreak in December 2013 to May 2015 using historic Twitter data. Although there have been several studies performed on Ebola tweet datasets, they were smaller and covered various periods of time. (4–10) My study has enabled synthesis of information from previous studies, and contributed new evidence. This is the largest and most complete dataset of more than 60 million Ebola related tweets. It included all Twitter data that met the filtering criteria, not just a sample. This enabled me to perform absolute measurements such as to measure the number of unique users participating in the conversation or to identify the most retweeted tweets. It was also the first study that included the tweets containing URLs to the articles that had the word “Ebola” in their title, description, or among their keywords, but not necessarily mentioning the word “Ebola” or one of the hashtags in the tweet text. This

approach allowed me to add approximately 7.5 million tweets to the dataset, which eventually comprised approximately 11.5% of the tweets in the final dataset. In addition to the metadata provided with tweets, I was able to use augmentation services of DataSift (data source), which provided additional data on tweet languages, based on the tweet text itself, rather than relying on the self-reported languages of Twitter users. Moreover, link augmentation enabled me to resolve URLs in tweets to their end points and analyse them. To my knowledge, this is the first study that has identified the Twitter accounts that achieved the highest level of engagement with other Twitter users and examined the place of governmental and non-governmental organisations among them. It was also the first to analyse the Ebola related Twitter activity of the WHO and CDC, to perform content analysis of their tweets, and to examine their engagement with other Twitter users related to the Ebola outbreak.

Moreover, this study contributes new methodologies for future research of the thematic Twitter datasets that correlate with news events. Firstly, I developed a way of studying a dataset covering a long period of time by breaking it down into smaller sub-periods based on the tweet frequency distribution. Further, I took an advantage of having a high prevalence of tweets containing URLs in order to identify the main topic of discussion on the days of interest.

The main limitation of this study is that despite the fact that this is the most complete Ebola tweet dataset studied up to date, it still has tweets missing for several reasons. Firstly, some tweets may be missing because not all Ebola-related tweets were covered by the filtering criteria. There might be more relevant key words and hashtags that were not included in the filtering criteria. In addition, it is not possible to include all tweets that may indirectly refer to Ebola. Secondly, filtering criteria were based on Latin script which excluded all tweets that used any other script for a key word or hashtags. In addition, only English language hashtags were included. This might have left out tweets that included

hashtags in any other languages that use Latin Script (such as French, Spanish, and Portuguese) and did not meet the rest of the filtering criteria. Due to that reason a part of the conversation from several affected countries such as Guinea or Spain may not be reflected in the dataset. However, in many cases tweets in these languages still contained the word or the hashtag “ebola” and were included in the study dataset. Thirdly, although the data provider attempted to deliver 100% of the data, due to unavoidable disruptions during the delivery process and other unforeseen issues, some data were not delivered and the dataset did not include the entirely comprehensive corpus of all the tweets that met the filtering criteria. However, the identified gaps did not cover more than two hours. Nevertheless, some additional data that were not identified can be absent. It was also not possible to verify the algorithms behind the services of the data provider due to the commercial nature of this information. Lastly, some of the tweets and accounts might have been deleted by the time I obtained the data for this project, making them not available within the Twitter historic data. As Twitter itself was not involved in this research, there was no opportunity to independently verify the quality of the dataset. Owing to the number of tweets missing in the dataset, all findings of absolute measurements (such as most retweeted tweets or numbers of followers) should be interpreted with caution. Some relative findings, such as proportions or temporal frequencies could also be slightly distorted. However, this unavoidable limitation does not make the results of this study invalid. Moreover, the findings of this study are only generalisable to tweets containing the word Ebola and hashtags of interest in Latin script.

The other limitation of the dataset is that it might have included some irrelevant tweets. Firstly, due to the fact that Ebola conversation was trending over certain periods of time, it is likely that Ebola related hashtags got “spammed”, as it has been reported for many other trending topics. (181) This means that “spammers” included these hashtags in their tweets in order to make them visible in search or trending topics. Secondly, link

augmentation also brought some irrelevant tweets, which did not refer to the actual Ebola story. It is not possible to determine the prevalence of these irrelevant tweets and how evenly they are distributed in the dataset, so it should be considered that these tweets could distort findings containing relative measurements.

Even though the Twitter population is global, it does not represent the general population due to the demographic and socioeconomic differences of Twitter users. Twitter usage has been shown to be larger in younger groups and minority ethnic populations. (182) Moreover, Twitter popularity and accessibility varies in different geolocations, in urban and rural areas. This is especially the case for those most affected by Ebola in West African countries. Some places have restricted access to the Internet. Moreover, in some countries, such as China, Twitter is blocked by the government. Therefore, the findings of this study apply only to the population of Twitter users.

There is also a limitation for the interpretation of the tweet geographical distribution analysis results, which only shows the geographical distribution of tweets that were posted from mobile devices or with the use of third-party applications and their authors actively enabled their geo-tagging. Therefore, the geography of tweets reported in this study was likely to be biased due to the fact that these are mainly affluent societies that access Twitter from mobile devices. The imputation of geo-data using existing computational algorithms could provide further insights on the geographical distribution of tweets in the dataset. Automated tweet aggregator accounts that posted a large volume of data may have indicated their precise location. This could introduce bias in geographical mapping of study sample tweets. The identification and exclusion of these accounts from the dataset may improve the geographical distribution analysis results.

While comparing the tweet frequency distribution with the Ebola epidemiological data and Ebola news events, the correspondence of tweet frequency with the news events might be due to chance, and there may be other reasons that could influence the tweet

frequency over the study period. Nevertheless, my findings are consistent with the other findings on this issue. Moreover, I did not adjust the temporal tweet frequency distribution to the increasing number of the reported Twitter users over time, as the tweet frequency distribution did not increase with time as the number of Twitter users did. (183) This should be taken into account by future studies who may want to compare the findings of my study with the results from other data sources.

The content analysis of accounts that achieved the highest level of Twitter engagement as well as the WHO and CDC tweets was conducted manually and only by one researcher, making it prone to interpretation bias. This can be improved in the future by involving a second researcher. However, the manual content analysis has demonstrated the highest level of accuracy in the previous studies of Twitter data. (114)

Although the acquired Twitter data sample on Ebola cannot be made available to other researchers according to Twitter policy, I have stated all the filtering criteria that I used for the study sample collection. This makes it possible for other researchers to obtain the same sample from Twitter for further investigation. Moreover, based on this dataset a public portal will be developed in collaboration with Splunk4Good programme. It will enable the public to explore other insights from these data through a set of online dashboards.

Due to the unusual nature of the term “Ebola” (this word does not have any other meaning⁵² and defines the name of the virus and the disease in most of the languages that use Latin script), the methodology for data filtering that was used in this study cannot always be applied in other studies that use Twitter data for health topic research. For example, it would be more difficult to identify a Twitter data sample about “stroke”, as the word has several meanings and various names in different languages, even among those using Latin script.

⁵² The exception is the name of the river in the Democratic Republic of Congo after which the virus takes its name.

Although all the presented limitations do not invalidate the findings of this research, the findings must be regarded as exploratory and speculative.

6.4. Wider implications of the study and lessons learned

The findings of this study contribute to the growing body of evidence about Ebola conversation on Twitter. I performed the study using Ebola tweets covering 528 days, which enables the synthesis of the existing evidence on the pattern of Twitter use by the general population over the period of the outbreak. It fills the gaps in evidence by including the time periods that have not been previously studied. It demonstrates the geographical distribution of the tweets posted in the different periods, showing which parts of the world were active in the conversation. It provides new evidence that the accounts of news organisations, humour accounts, Twitter celebrities, and celebrities achieved the highest level of Twitter engagement with other users over the period of the Ebola outbreak and demonstrates that the general Twitter population turned to the accounts of public health authorities to seek public health information. It also shows that when they provided the information, Twitter users actively engaged with it, by reading it, distributing to a wider audience, mentioning them and sending replies.

Policy implications

Through this investigation, I have demonstrated that Twitter is an important tool of communication during public health emergencies. The general Twitter population used it to share and seek information and turned to the accounts of global health organisations for expertise. Therefore, health authorities such as the WHO and CDC need to keep building their presence on Twitter in order to track misinformation and provide accurate information. Global health organisations used Twitter to provide accurate information and engage with their Twitter audience, answering their questions and understanding their

concerns to inform overall communication strategy. Although the Internet is not universally available, and the access to Twitter and other social media platforms is still limited, more and more areas in the world are achieving broad Internet coverage, and the technology enabling the access to social media is becoming more affordable. Therefore, global health organisations must be ready to communicate with the general public through the social media platforms they use, such as Twitter. However, it is also important for them to build their presence on many other social media platforms, such as Facebook or Instagram. Popularity of different social media platforms is constantly changing, so it is important for these organisations to track this and be active where the people are active.

Moreover, this study demonstrates that it is important for public health organisations to identify accounts that achieve high level of Twitter engagement during public health emergencies due to several reasons. Firstly, it is important to monitor and be aware of the content posted by these accounts in order to detect and address any inaccurate information. For example, the popularity of humour accounts and the accounts of Twitter celebrities during the West African Ebola outbreak was beyond the control of public health organisations and could have potentially led to dissemination of misinformation and decreased the credibility of public health authorities. On the other hand, public health organisations could consider collaboration with these accounts in order to be retweeted by them or to include accurate public health content in their messages. Furthermore, the understanding of what content features lead to a wider Twitter engagement may help public health organisations to strategically incorporate them in their messages. For example, humour accounts demonstrated a high level of Twitter engagement amongst the general Twitter population during the West African Ebola outbreak, due to them sharing humorous content. This fact would demonstrate that during periods of public panic and anxiety people pay attention to and share humorous content. If so, public health

organisations should be aware of this issue and strategically include humorous and emotional content in their tweets for a wider dissemination of accurate information.

The volume of data in the dataset required the use of Big Data analysis tools and this study illustrates the utility of applications, such as Splunk, for the analysis of really large datasets in a manageable and quick way. Moreover, these applications are flexible and enable the change of filters and search terms to track real time changes during the events of interest. The study demonstrates a variety of methods that can be applied to a Twitter dataset that can be used by different actors participating in an emergency response, including public health authorities, in order to answer a range of questions. For example, global health authorities, such as WHO, may identify locations where the majority of tweets come from in order to understand which parts of the world express a greater concern about a situation and identify a target audience for their messages, as well as events which rise public concern. They also may quickly identify what kind of accounts achieve the highest level of engagement with the Twitter population and examine their own position relatively to them. This information could enable health authorities to track the information provided by these accounts, and if needed, addressing it. It might also help to identify the accounts with which they may consider collaboration for the information dissemination. Moreover, the use of Big Data analytic tools provides opportunities for public health organisations to assess their own performance on social media during public health emergencies in real time and compare it with the performance of other Twitter accounts of interest.

However, the difficulties in obtaining, cleaning, and managing Twitter data should be considered by health authorities, as well as limitations in interpretation of results. Since the full Twitter data are not free, perhaps these organisations could directly collaborate with the platform in obtaining them, or use real-time Twitter data to estimate relative measurements. Algorithms for cleaning and managing the data could be established in

order to obtain the results that can inform organisations' communication strategies. A variety of visualisation tools make the results easily presentable in order to be quickly considered and understood by a decision-making body.

Future research implications

Future researchers should focus on improving the quality of the filtering criteria for the Ebola-related Twitter dataset. Its completeness may be improved by the inclusion of non-Latin script tweets as well as popular hashtags in other languages that use Latin script. There is also a need to develop further algorithms that can identify tweets indirectly referring to Ebola. Moreover, irrelevant tweets that were included in the dataset, such as tweets with spammed hashtags or posted by automated tweet aggregators, should be identified and excluded.

The accuracy and limitations of URL use for the identification of main discussion topics on Twitter driven by media coverage can be further explored. The minimum prevalence of URLs and retweets in a data sample and the maximum time period that this sample can cover in order to provide accurate insights should be established.

There is also a need to further explore who the Twitter users tweeting from West African countries during the Ebola outbreak were, and whether they were citizens of the affected countries or foreigners, such as international workers assisting in the outbreak. This will help to understand whether Ebola-related information exchange on Twitter involved populations most affected by the outbreak.

Further work is needed in order to investigate what role different types of accounts play in information dissemination on Twitter during public health emergencies and what strategies they use to achieve a higher level of engagement with the general Twitter population. A separate analysis can investigate what news agencies achieved the highest level of Twitter engagement around the world and in different geographical locations, what the driving forces behind that were, and how accurate the information provided by them

was. Further research is also needed to establish the role humour accounts and humorous content play on Twitter during public health emergencies and to identify the type of content humour accounts actually tweeted during the West African Ebola outbreak. An examination of the reasons why the general public actively engage with these accounts, especially during the periods when high levels of anxiety are expressed, might help to inform social media strategies of public health organisations. The accounts of Twitter celebrities can also be further explored to understand who the people or organisations owning them are, how accurate Ebola-related information shared by them was, and what kinds of content led to the broad dissemination of their tweets. Finally, the way other health organisations that achieved a high level of Twitter engagement (such as UNICEF and UNMEER) used Twitter can be explored and compared to the way the WHO and CDC did. These findings can be further compared with the actual social media strategies used by these organisations in order to provide more specific recommendations to them. Future research may examine the content of tweets posted by health authorities to understand which factors led to more retweeting during the Ebola outbreak. Moreover, the relationship between different types of accounts that achieved the highest level of engagement can be investigated further in order to establish whether their popularity in different periods of the outbreak was independent from each other or if any correlation may be observed.

Moreover, the findings of this research might be further supported by qualitative methods of research such as interviews or surveys of Twitter users who tweeted about Ebola, in order to investigate their motivations to tweet and retweet the information about Ebola, to engage with the accounts of news organisations, humorous accounts, Twitter celebrities, and celebrities. It might also help to identify the impact of fundraising or public health messages posted by these accounts and help to determine whether the information provided by the WHO and CDC had an impact and stimulated positive behaviour change.

6.5. Conclusion

I have used novel Big Data analysis tools to demonstrate that Twitter activity during the West African Ebola outbreak was more likely to coincide with news events rather than disease dynamics. I have shown that during the Ebola conversation on Twitter, its users were more likely to retweet each other sharing news and information than during the general Twitter conversation. However, the direct engagement of Twitter users with each other was relatively low. The accounts of news organisations, humour accounts, accounts of Twitter celebrities, and celebrities were among those who achieved the highest level of engagement with the general Twitter population. Accounts of such governmental health organisations as the WHO and CDC also ranked high in engagement levels. It is important for health organisations to consider the fact that during public health emergencies the general Twitter public receives a high volume of information from news organisations, humour accounts, Twitter celebrities, and celebrities. Therefore, it is important to monitor how accurate this information is and consider collaboration with these accounts in order to disseminate important information. Public health organisation could also consider how to strategically use humorous and emotional content in their tweets to stimulate a wider dissemination of public health information. The study demonstrates the utility of Big Data analytics tools, such as Splunk, for the analysis of Twitter data in order to quickly yield insights.

Twitter was an important communication tool for global health organisations and was widely used by the WHO and CDC in relation to the West African Ebola outbreak both for one-way and two-way communication. Health authorities would benefit from building their Twitter presence because, as has been demonstrated in the example with the Ebola outbreak, people turned to the accounts of these organisations in order to ask questions and obtain accurate public health information in public health emergencies. It is evident

from my findings that CDC was more responsive to their public than WHO, in that the CDC was responding to questions and posting required information in a timely manner. Health authorities could consider the use of Big Data analytics tools in order to quickly gain insights from social media data to inform their communications strategies.

Given the limitations of the current dataset, such as the use of Latin script in the filtering criteria and the presence of tweets with spammed hashtags, further research which includes tweets in other languages and identifies a way of filtering out spammed tweets would be beneficial. Moreover, in order to get a clearer picture of geographical distribution of tweets, future studies could compare the results of this study with the analysis of tweets with the imputed geo-data. Further investigation of the accounts that achieved a high level of Twitter engagement, the reasons for that, and the relationship between them would help better inform communication strategies of public health organisations. If the findings of this study were to be compared with the results from different data sources, they should be adjusted to the increasing number of Twitter users over the study period.

Acknowledgements

I would like to thank:

- The ExxonMobil Foundation and Dr. Proochista Ariana, for making this possible.
- My supervisors Dr. Francois Van Loggerenberg and Dr. Proochista Ariana, for their feedback and suggestions for improvement and for their patient guidance throughout the project and writing process.
- My colleagues Dr. Pouria Amirian, Dr. Arthur Thomas, and Dr. Scott Hale for their methodology advice and technical assistance.
- Colleagues at the Global Health Network, for their warm hospitality and support throughout the year.
- Splunk4Good corporate social responsibility programme of Splunk Inc. and Splunk team members, especially Corey Marshall and Michael Wilde, for collaboration, training, constant assistance throughout the project and provided financial support for the data acquisition.
- My Oxford friends for their support and friendship. They made this year so special and made it an unforgettable experience.
- My friend Guy Lomax, for the proof reading of this thesis.
- My parents, for their constant encouragement, love and moral support. To them I dedicate this thesis.

References

1. Company | About [Internet]. Twitter About. [cited 2015 Sep 27]. Available from: <https://about.twitter.com/company>
2. Kurka DB, Godoy A, Zuben FJ Von. Online social network analysis: a survey of research applications in computer science. ArXiv150405655 Phys [Internet]. 2015 Apr 22 [cited 2015 May 8]; Available from: <http://arxiv.org/abs/1504.05655>
3. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. *N Engl J Med*. 2011 Jul 28;365(4):289–91.
4. Carter M. How Twitter may have helped Nigeria contain Ebola. *BMJ*. 2014 Nov 19;349:g6946.
5. Fung IC-H, Tse ZTH, Cheung C-N, Miu AS, Fu K-W. Ebola and the social media. *The Lancet*. 2014 Dec;384(9961):2207.
6. Kalyanam J, Sumithra Velupillai, Doan S, Conway M, Lanckriet G. Facts and fabrications about Ebola: A Twitter based study. ArXiv150802079 Cs [Internet]. 2015 Aug 9 [cited 2015 Sep 5]; Available from: <http://arxiv.org/abs/1508.02079>
7. Lee TM. Ebola on social media shows some revealing insights [Internet]. Symplur. [cited 2015 May 24]. Available from: <http://www.symplur.com/blog/ebola-social-media-first-look-facts/>
8. Luckerson V. Fear, misinformation, and social media complicate Ebola fight. *Time* [Internet]. 2014 Oct 8 [cited 2015 Feb 3]; Available from: <http://time.com/3479254/ebola-social-media/>
9. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *Am J Infect Control*. 2015 Jun;43(6):563–71.
10. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ*. 2014 Oct 14;349:g6178.
11. Cohen H. Social media definitions [Internet]. Heidi Cohen. [cited 2015 Mar 22]. Available from: <http://heidicohen.com/social-media-definition/>
12. Whiting A, Williams D. Why people use social media: a uses and gratifications approach. *Qual Mark Res Int J*. 2013 Aug 30;16(4):362–9.
13. Number of worldwide social network users 2010-2018 | Statistic| [Figure], Number of social network users worldwide from 2010 to 2018 (in billions). [Internet]. Statista. 2015 [cited 2015 Mar 22]. Available from: <http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
14. Social networks: global sites ranked by users 2015 | Statistic| [Figure], Social networks: global sites ranked by users 2015 [Internet]. Statista. 2015 [cited 2015 Mar 22].

Available from: <http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

15. Scaling the Facebook data warehouse to 300 PB [Internet]. Facebook Code. [cited 2015 Mar 23]. Available from: <https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/>
16. Focusing on Efficiency | Facebook Newsroom [Internet]. [cited 2015 Mar 23]. Available from: <http://newsroom.fb.com/news/2013/09/focusing-on-efficiency/>
17. Here's to eight great years [Internet]. Official YouTube Blog. [cited 2015 Mar 23]. Available from: <http://youtube-global.blogspot.com/2013/05/heres-to-eight-great-years.html>
18. Mayer-Schonberger V, Cukier K. Big Data: a revolution that will transform how we live, work and think. Hachette UK; 2013. 239 p.
19. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009 Feb 19;457(7232):1012–4.
20. Google Flu Trends. How does this work? [Internet]. [cited 2015 Mar 24]. Available from: <http://www.google.org/flutrends/about/how.html>
21. Google Flu Trends. How does this work? [Figure], Correlation between Google Flu Trends and CDC surveillance data for the USA (2004-2014) and National Institute of Infectious Diseases for Japan (2004-2009). [Internet]. 2015 [cited 2015 Mar 24]. Available from: <http://www.google.org/flutrends/about/how.html>
22. Kaplan AM, Haenlein M. The early bird catches the news: nine things you should know about micro-blogging. *Bus Horiz*. 2011 Mar;54(2):105–13.
23. Twitter: number of monthly active users 2014 | Statistic| [Figure], Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2015 (in millions). [Internet]. Statista. 2015 [cited 2015 Mar 22]. Available from: <http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>
24. Twitter Help Center. FAQs about verified accounts. [Internet]. [cited 2015 Feb 9]. Available from: <https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts>
25. Liu Y, Kliman-Silver C, Mislove A. The Tweets they are a-changin': evolution of Twitter users and behavior. In: Eighth International AAAI Conference on Weblogs and Social Media [Internet]. 2014 [cited 2015 Sep 4]. p. 305–14. Available from: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8043>
26. The anatomy of a Tweet: metadata on Twitter – Slaw| [Figure], Map of Twitter status object (tweet metadata) as of April 2010 [Internet]. 2010 [cited 2015 May 25]. Available from: <http://www.slaw.ca/2011/11/17/the-anatomy-of-a-tweet-metadata-on-twitter/>

27. Duggan M, Ellison NB, Lampe C, Am, Lenhart a, Madden M. Social media update 2014 [Internet]. Pew Research Center's Internet & American Life Project. [cited 2015 May 7]. Available from: <http://www.pewinternet.org/2015/01/09/social-media-update-2014/>
28. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A. The Twitter of babel: mapping world languages through microblogging platforms. PLoS ONE. 2013 Apr 18;8(4):e61981.
29. Summary by language size [Internet]. Ethnologue. [cited 2015 May 7]. Available from: <http://www.ethnologue.com/statistics/size>
30. The countries that block Twitter, Facebook and YouTube [MAP] [Internet]. [cited 2015 May 7]. Available from: <http://www.adweek.com/socialtimes/countries-block-twitter/497573>
31. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A. The Twitter of babel: mapping world languages through microblogging platforms. PLoS ONE. 2013 Apr 18;8(4):e61981. Figure 5, Languages by number of users; eb1981.
32. Liu Y, Kliman-Silver C, Mislove A. The Tweets they are a-changin': evolution of Twitter users and behavior. In: Eighth International AAAI Conference on Weblogs and Social Media [Internet]. 2014 [cited 2015 Sep 4]. p. 305–14. Figure 6, The percentage of users self – reporting the six most popular languages over time. We plot English separately in the top graph in order to increase the readability; note that the scale is different between the two graphs. English shows a mostly linear decrease from 83% in January 2010 to 52% in December 2013; p. 309. Available from: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8043>
33. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A. The Twitter of babel: mapping world languages through microblogging platforms. PLoS ONE. 2013 Apr 18;8(4):e61981. Figure 7, Language share of the most active countries.; eb1981.
34. Twitter Help Center. Adding your location to a Tweet. [Internet]. [cited 2015 Feb 9]. Available from: <https://support.twitter.com/articles/122236>
35. FAQs about adding location to your Tweets [Internet]. Twitter Help Center. [cited 2015 Dec 23]. Available from: <https://support.twitter.com/articles/78525>
36. Leetaru K, Wang S, Cao G, Padmanabhan A, Shook E. Mapping the global Twitter heartbeat: the geography of Twitter. First Monday [Internet]. 2013 Apr 22 [cited 2015 Mar 29];18(5). Available from: <http://firstmonday.org/ojs/index.php/fm/article/view/4366>
37. Leetaru K, Wang S, Cao G, Padmanabhan A, Shook E. Mapping the global Twitter heartbeat: the geography of Twitter. First Monday. 2013 Apr 22;18(5):Figure 5, All Exact Location coordinates in the Twitter Decahose 23 October 2012 to 30 November 2012.
38. Heldman AB, Schindelar J, Weaver J. Social media engagement and public health communication: implications for public health organizations being truly" social. Public Health Rev. 2013;35(1):1–1.
39. Rogers R. Debanalizing Twitter: the transformation of an object of study. In: Proceedings of the 5th Annual ACM Web Science Conference [Internet]. New York, NY, USA:

ACM; 2013 [cited 2015 May 6]. p. 356–65. Available from:
<http://doi.acm.org/10.1145/2464464.2464511>

40. Finfgeld-Connett D. Twitter and health science research. *West J Nurs Res*. 2014 Dec 26;1269–83.
41. Health fact sheet [Internet]. Pew Research Center’s Internet & American Life Project. [cited 2015 May 11]. Available from: <http://www.pewinternet.org/fact-sheets/health-fact-sheet/>
42. PWC Health Research Institute. Social media “likes” healthcare. From marketing to social business. [Internet]. 2012 [cited 2015 Nov 5]. Available from: <http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml>
43. Shane D. Company use of Twitter is significantly increasing [Internet]. *Small Business Trends*. [cited 2015 May 18]. Available from: <http://smallbiztrends.com/2012/09/company-use-of-twitter-is-significantly-increasing.html>
44. Social Times. 7 ways businesses can use Twitter [Internet]. [cited 2015 May 18]. Available from: <http://www.adweek.com/socialtimes/7-ways-businesses-can-use-twitter/445478>
45. Lovejoy K, Saxton GD. Information, community, and action: how nonprofit organizations use social media [Internet]. Rochester, NY: Social Science Research Network; 2011 Jun [cited 2015 May 18]. Report No.: ID 2039815. Available from: <http://papers.ssrn.com/abstract=2039815>
46. Bermúdez-Tamayo C, Alba-Ruiz R, Jiménez-Pernett J, García Gutiérrez J-F, Traver-Salcedo V, Yubraham-Sánchez D. Use of social media by Spanish hospitals: perceptions, difficulties, and success factors. *Telemed E-Health*. 2013 Jan 31;19(2):137–45.
47. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media’s capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. *Health Promot Pract*. 2013 Mar;14(2):157–62.
48. Carrillo-Larco RM. Social networks and public health: use of Twitter by ministries of health. *Int J Public Health*. 2012 Aug;57(4):755–6.
49. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. *BMC Public Health*. 2012;12:242.
50. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. *Am J Public Health*. 2013 Sep;103(9):1700–7.
51. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of Twitter among local health departments: an analysis of information sharing, engagement, and action. *J Med Internet Res* [Internet]. 2013 Aug 19 [cited 2015 May 18];15(8). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758023/>

52. Harris JK, Mueller NL, Snider D, Haire-Joshu D. Local health department use of Twitter to disseminate diabetes information, United States. *Prev Chronic Dis* [Internet]. 2013 May 2 [cited 2015 May 18];10. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3652718/>
53. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The Use of Social Media by State Tobacco Control Programs to Promote Smoking Cessation: A Cross-Sectional Study. *J Med Internet Res* [Internet]. 2014 Jul 10 [cited 2015 May 19];16(7). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4115651/>
54. Villa Hernández TE, Ocampo MA, Cicero Sabido R. [Internet social networks tobacco outreach: the clinical experience against tobacco of the General Hospital of Mexico "Eduardo Liceaga"]. *Gac Médica México*. 2012 Oct;148(5):438–47.
55. Rocheleau M, Sadasivam RS, Baquis K, Stahl H, Kinney RL, Pagoto SL, et al. An observational study of social and emotional support in smoking cessation Twitter accounts: content analysis of Tweets. *J Med Internet Res*. 2015 Jan 14;17(1):e18.
56. Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S, et al. A systematic examination of the use of online social networking sites for sexual health promotion. *BMC Public Health*. 2011 Jul 21;11:583.
57. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J, et al. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. *MMWR Morb Mortal Wkly Rep*. 2014 Aug 15;63(32):681–5.
58. Prabhu V, Lee T, Loeb S, Holmes JH, Gold HT, Lepor H, et al. Twitter response to the United States preventive services task force recommendations against screening with prostate-specific antigen. *BJU Int*. 2014 Mar 25;
59. King D, Ramirez-Cano D, Greaves F, Vlaev I, Beales S, Darzi A. Twitter and the health reforms in the English National Health Service. *Health Policy*. 2013 May;110(2-3):291–7.
60. WHO | Mixed uptake of social media among public health specialists [Internet]. WHO. [cited 2015 Sep 6]. Available from: <http://www.who.int/bulletin/volumes/89/11/11-031111/en/>
61. WHO finds social media indispensable in managing global health crises [Internet]. The Huffington Post. [cited 2015 Sep 6]. Available from: http://www.huffingtonpost.com/david-j-olson/who-social-media_b_1530016.html
62. U.S. Centers for Disease Control and Prevention. Social media at CDC [Internet]. Available from: <http://www.cdc.gov/socialmedia/tools/twitter.html>
63. U.S. Centers for Disease Control and Prevention. The Health Communicator's Social Media Toolkit [Internet]. 2011. Available from: <http://www.cdc.gov/socialmedia/tools/guidelines/socialmediatoolkit.html>
64. Quincey E de, Kostkova P. Early warning and outbreak detection using social networking websites: the potential of Twitter. In: Kostkova P, editor. *Electronic Healthcare* [Internet]. Springer Berlin Heidelberg; 2010 [cited 2015 May 12]. p. 21–4. Available from:

http://link.springer.com/chapter/10.1007/978-3-642-11745-9_4

65. Kostkova P, Szomszor M, St. Louis C. #Swineflu: The Use of Twitter As an Early Warning and Risk Communication Tool in the 2009 Swine Flu Pandemic. *ACM Trans Manage Inf Syst.* 2014 Jul;5(2):8:1–8:25.
66. Wasantha Jayawardene AY. Role of social media in early warning of Norovirus outbreaks: a longitudinal Twitter-based infoveillance. Paper presented at: *WORLDCOMP'13 – 9th International Conference on Data Mining (DMIN'13)*. In Las Vegas Nevada, USA; 2013.
67. Gomide J, Veloso A, Meira W, Almeida V, Benevenuto F, Ferraz F, et al. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In Koblenz, Germany; 2011 [cited 2015 May 14]. p. 1–8. Available from: http://www.websci11.org/fileadmin/websci/Papers/92_paper.pdf
68. Kriek M, Dreesman J, Otrusina L, Denecke K. A new age of public health: Identifying disease outbreaks by analyzing tweets. In: *Proceedings of Health WebScience Workshop, ACM Web Science Conference.* 2011.
69. Achrekar H, Gandhe A, Lazarus R, Yu S-H, Liu B. Predicting Flu Trends using Twitter data. In: *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).* 2011. p. 702–7.
70. Aramaki E, Maskawa S, Morita M. Twitter catches the Flu: detecting influenza epidemics using Twitter. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing [Internet].* Stroudsburg, PA, USA: Association for Computational Linguistics; 2011 [cited 2015 May 12]. p. 1568–76. Available from: <http://dl.acm.org/citation.cfm?id=2145432.2145600>
71. Kim E-K, Seok JH, Oh JS, Lee HW, Kim KH. Use of hangeul Twitter to track and predict human Influenza infection. *PLoS ONE [Internet].* 2013 Jul 24 [cited 2015 May 13];8(7). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3722273/>
72. Broniatowski DA, Paul MJ, Dredze M. National and local Influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS ONE [Internet].* 2013 Dec 9 [cited 2015 May 13];8(12). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3857320/>
73. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. In: *Proceedings of the First Workshop on Social Media Analytics [Internet].* New York, NY, USA: ACM; 2010 [cited 2015 May 12]. p. 115–22. Available from: <http://doi.acm.org/10.1145/1964858.1964874>
74. Paul MJ, Dredze M, Broniatowski D. Twitter improves Influenza forecasting. *PLoS Curr [Internet].* 2014 [cited 2015 May 14]; Available from: <http://currents.plos.org/outbreaks/?p=39911>
75. Collier N, Son NT, Nguyen NM. OMG U got flu? Analysis of shared health messages for bio-surveillance. *J Biomed Semant.* 2011 Oct 6;2(Suppl 5):S9.

76. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naïve language. *PLoS ONE* [Internet]. 2013 Dec 4 [cited 2015 May 14];8(12). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3853203/>
77. Lamos V, Bie TD, Cristianini N. Flu detector - tracking epidemics on Twitter. In: Balcázar JL, Bonchi F, Gionis A, Sebag M, editors. *Machine Learning and Knowledge Discovery in Databases* [Internet]. Springer Berlin Heidelberg; 2010 [cited 2015 May 14]. p. 599–602. Available from: http://link.springer.com/chapter/10.1007/978-3-642-15939-8_42
78. Lamos V, Cristianini N. Tracking the flu pandemic by monitoring the social web. In: *2010 2nd International Workshop on Cognitive Information Processing (CIP)*. 2010. p. 411–6.
79. Lee TM. The life cycle of Ebola on Twitter [Internet]. *Symplur*. [cited 2015 May 24]. Available from: <http://www.symplur.com/blog/the-life-cycle-of-ebola-on-twitter/>
80. Jung J-Y, Moro M. Multi-level functionality of social media in the aftermath of the Great East Japan Earthquake. *Disasters*. 2014 Jul;38 Suppl 2:S123–43.
81. Cates AL, Arnold BW, Cooper GP, Yeager V, Stake J, Ali M, et al. Impact of dual-polarization radar technology and Twitter on the Hattiesburg, Mississippi tornado. *Disaster Med Public Health Prep*. 2013 Dec;7(6):585–92.
82. Genes N, Chary M, Chason K. Analysis of Twitter users' sharing of official New York Storm response messages. *Med* 20. 2014 Mar 20;3(1):e1.
83. Tobias E. Using Twitter and other social media platforms to provide situational awareness during an incident. *J Bus Contin Emerg Plan*. 2011 Oct;5(3):208–23.
84. Splunk4Good presents at the White House! #FEMAThinkTank wrapup | Splunk Blogs [Internet]. [cited 2015 Sep 26]. Available from: <http://blogs.splunk.com/2013/02/19/splunk4good-presents-at-the-white-house-femathinktank-wrapup/>
85. Feldmann H, Geisbert TW. Ebola haemorrhagic fever. *The Lancet*. 2011 Mar;377(9768):849–62.
86. WHO | Ebola virus disease [Internet]. WHO. [cited 2015 May 24]. Available from: <http://www.who.int/mediacentre/factsheets/fs103/en/>
87. WHO | Ground zero in Guinea: the outbreak smoulders – undetected – for more than 3 months [Internet]. WHO. [cited 2015 May 26]. Available from: <http://www.who.int/csr/disease/ebola/ebola-6-months/guinea/en/>
88. WHO | Statement on the 1st meeting of the IHR Emergency Committee on the 2014 Ebola outbreak in West Africa [Internet]. WHO. [cited 2015 May 24]. Available from: <http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/>
89. World Health Organization. Ebola situation report - 23 September 2015 [Internet]. [cited 2015 Sep 25]. Available from: <http://apps.who.int/ebola/current-situation/ebola->

situation-report-2-september-2015

90. WHO | Ebola transmission in Liberia over. Nation enters 90-day intensive surveillance period [Internet]. WHO. [cited 2015 Sep 5]. Available from: <http://www.who.int/mediacentre/news/statements/2015/ebola-transmission-over-liberia/en/>
91. WHO | World on the verge of an effective Ebola vaccine [Internet]. WHO. [cited 2015 Sep 5]. Available from: <http://www.who.int/mediacentre/news/releases/2015/effective-ebola-vaccine/en/>
92. World Health Organization. Ebola situation report - 23 September 2015. Figure 1, Figure 1: Confirmed, probable, and suspected EVD cases worldwide. [Internet]. Figure 1, Figure 1: Confirmed, probable, and suspected EVD cases worldwide. [cited 2015 Sep 25]. Available from: <http://apps.who.int/ebola/current-situation/ebola-situation-report-2-september-2015>
93. Healthmap. 2014 Ebola outbreaks. [Internet]. [cited 2015 May 9]. Available from: <http://www.healthmap.org/ebola/#timeline>
94. African digital statistics 2014 [Internet]. Lagos. [cited 2015 May 24]. Available from: <http://socialmediaweek.org/lagos/2014/02/03/african-digital-statistics-2014/>
95. Social media in Nigeria: state of the art, and 3 marketing best practices | Digital in the round [Internet]. [cited 2015 Sep 14]. Available from: <http://www.digitalintheround.com/social-media-marketing-nigeria/>
96. Butler D. Largest ever Ebola outbreak is not a global threat. Nature [Internet]. 2014 Jul 30 [cited 2015 Feb 3]; Available from: http://www.nature.com/news/largest-ever-ebola-outbreak-is-not-a-global-threat-1.15640?WT.mc_id=FBK_NatureNews
97. Mitman G. Ebola in a stew of fear. N Engl J Med. 2014 Nov 6;371(19):1763–5.
98. Rodriguez-Morales AJ, Castañeda-Hernández DM, McGregor A. What makes people talk about Ebola on social media? A retrospective analysis of Twitter use. Travel Med Infect Dis. 2015 Feb;13(1):100–1.
99. Luckerson V. Watch how word of Ebola exploded in America. Time [Internet]. 2014 Oct 7 [cited 2015 Sep 5]; Available from: <http://time.com/3478452/ebola-twitter/>
100. Report: Ebola coverage on TV news plummeted after midterms [Internet]. Media Matters for America. [cited 2015 Sep 5]. Available from: <http://mediamatters.org/research/2014/11/19/report-ebola-coverage-on-tv-news-plummeted-afte/201619>
101. The Lancet. The medium and the message of Ebola. The Lancet. 2014 Nov;384(9955):1641.
102. Towers S, Afzal S, Bernal G, Bliss N, Brown S, Espinoza B, et al. Mass media and the contagion of fear: the case of Ebola in America. PLoS ONE. 2015 Jun 11;10(6):e0129179.

103. Fung IC-H, Tse ZTH, Cheung C-N, Miu AS, Fu K-W. Ebola and the social media. *The Lancet*. 2014 Dec;384(9961):2207. Figure, Temporal trends on Twitter and Google about Ebola and influenza (flu) before, during, and after Ebola cases in the USA, September to November, 2014, p. 2207.
104. International Telecommunication Union. Percentage of individuals using the Internet 2000-2012 [Internet]. [cited 2015 May 20]. Available from: http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/Individuals_Internet_2000-2012.xls
105. Zawacki K. Ebola tweets are missing the target [Internet]. *The Daily Beast*. 2014 [cited 2015 Feb 3]. Available from: <http://www.thedailybeast.com/articles/2014/08/03/ebola-tweets-are-missing-the-target.html>
106. Smailhodvic A, Andrew K, Hahn L, Womble PC, Webb C. Sample NLPDE and NLODE social-media modeling of information transmission for infectious diseases:case study Ebola. *ArXiv150100198 Phys* [Internet]. 2014 Dec 31 [cited 2015 Sep 5]; Available from: <http://arxiv.org/abs/1501.00198>
107. WHO | Ebola: Experimental therapies and rumoured remedies [Internet]. WHO. [cited 2015 May 24]. Available from: <http://www.who.int/mediacentre/news/ebola/15-august-2014/en/>
108. Press Announcements - FDA warns consumers about fraudulent Ebola treatment products [Internet]. [cited 2015 May 24]. Available from: <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm410086.htm>
109. Rolison JJ, Hanoch Y. Knowledge and risk perceptions of the Ebola virus in the United States. *Prev Med Rep*. 2015;2:262–4.
110. Schindelar J. University of Oxford - Ebola on Twitter project. [online]. E-mail to Tatiana Vorovchenko (tatiana.vorovchenko@gtc.ox.ac.uk) 2015 Aug 08 [cited 2015 Sept 06].
111. Goff DA, Kullar R, Newland JG. Review of Twitter for infectious diseases clinicians: useful or a waste of time? *Clin Infect Dis*. 2015 May 15;60(10):1533–40.
112. The search API [Internet]. Twitter Developers. [cited 2015 Mar 11]. Available from: <https://dev.twitter.com/rest/public/search>
113. The streaming APIs [Internet]. Twitter Developers. [cited 2015 Feb 20]. Available from: <https://dev.twitter.com/streaming/overview>
114. Bruns A, Stieglitz S. Towards more systematic Twitter analysis: metrics for tweeting activities. *Int J Soc Res Methodol*. 2013 Mar;16(2):91–108.
115. Kim AE, Hansen HM, Murphy J, Richards AK, Duke J, Allen JA. Methodological considerations in analyzing Twitter data. *JNCI Monogr*. 2013 Dec 1;2013(47):140–6.
116. Christy JM. Demographics Pro information for University of Oxford [online]. E-mail to Tatiana Vorovchenko (tatiana.vorovchenko@gtc.ox.ac.uk) 2015 Apr 21 [cited 2015 May

25]. 2015.

117. Nwosu AC, Debattista M, Rooney C, Mason S. Social media and palliative medicine: a retrospective 2-year analysis of global Twitter data to evaluate the use of technology to communicate about issues at the end of life. *BMJ Support Palliat Care*. 2015 Jun;5(2):207–12.

118. Ramagopalan S, Wasiak R, Cox AP. Using Twitter to investigate opinions about multiple sclerosis treatments: a descriptive, exploratory study. *F1000Research*. 2014;3:216.

119. Dyar OJ, Castro-Sánchez E, Holmes AH. What makes people talk about antibiotics on social media? A retrospective analysis of Twitter use. *J Antimicrob Chemother*. 2014 Sep;69(9):2568–72.

120. Cavazos-Rehg P, Krauss M, Gruzca R, Bierut L. Characterizing the followers and Tweets of a marijuana-focused Twitter handle. *J Med Internet Res [Internet]*. 2014 Jun 27 [cited 2015 Feb 19];16(6). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4090385/>

121. Harris JK, Moreland-Russell S, Tabak RG, Ruhr LR, Maier RC. Communication about childhood obesity on Twitter. *Am J Public Health*. 2014 Jul;104(7):e62–9.

122. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Gruzca RA, Bierut LJ. Twitter chatter about marijuana. *J Adolesc Health Off Publ Soc Adolesc Med*. 2015 Feb;56(2):139–45.

123. Powell RJ, Kim AE, Richards A, et al. Quantity versus quality: the impact of sampling tweets on health care reform opinions. In: *Conference of the Midwest Association for Public Opinion Research*. Chicago, IL.; 2012.

124. Yoon S, Elhadad N, Bakken S. A practical approach for content mining of Tweets. *Am J Prev Med*. 2013 Jul;45(1):122–9.

125. Chen P-C, Lee H-M, Tyan H-R, Wu J-S, Wei T-E. Detecting spam on Twitter via message-passing based on Retweet-relation. In: Cheng S-M, Day M-Y, editors. *Technologies and Applications of Artificial Intelligence [Internet]*. Springer International Publishing; 2014 [cited 2015 Mar 16]. p. 56–65. Available from: http://link.springer.com/chapter/10.1007/978-3-319-13987-6_6

126. Guo D, Chen C. Detecting non-personal and spam users on geo-tagged Twitter network. *Trans GIS*. 2014 Jun 1;18(3):370–84.

127. ETL - Extract, Transform, Load [Internet]. *Webopedia*. 2015 [cited 2015 Oct 9]. Available from: <http://www.webopedia.com/TERM/E/ETL.html>

128. Murphy J. Relatively normal? Adjusting Tweet volume to account for Twitter's rise in popularity | *SurveyPost [Internet]*. [cited 2015 May 25]. Available from: <https://blogs.rti.org/surveypost/2012/04/27/relatively-normal-adjusting-tweet-volume-to-account-for-twitthers-rise-in-popularity/>

129. Ghosh D (Debs), Guha R. What are we “tweeting” about obesity? Mapping tweets with Topic Modeling and Geographic Information System. *Cartogr Geogr Inf Sci*.

2013;40(2):90–102.

130. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes MD. “Right time, right place” health communication on Twitter: value and accuracy of location information. *J Med Internet Res*. 2012 Nov 15;14(6):e156.

131. Dredze M, Paul MJ, Bergsma S, Tran H. Carmen: a Twitter geolocation system with applications to public health.

132. Suh B, Hong L, Pirolli P, Chi EH. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *IEEE*; 2010 [cited 2015 Sep 4]. p. 177–84. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5590452>

133. Linshi J. Twitter: no, 23 Million of our active users are not actually bots. *Time* [Internet]. 2014 Aug 12 [cited 2015 Sep 3]; Available from: <http://time.com/3103867/twitter-bots/>

134. Stringhini G, Egele M, Kruegel C, Vigna G. Poultry markets: on the underground economy of twitter followers. In *ACM Press*; 2012 [cited 2015 Sep 4]. p. 1. Available from: <http://dl.acm.org/citation.cfm?doid=2342549.2342551>

135. Bruns A, Stieglitz S. Quantitative approaches to comparing communication patterns on Twitter. *J Technol Hum Serv*. 2012 Jul 1;30(3-4):160–85.

136. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In: 2010 43rd Hawaii International Conference on System Sciences (HICSS). 2010. p. 1–10.

137. What fuels a Tweet’s engagement? [Internet]. *Twitter Blogs*. [cited 2015 Sep 3]. Available from: <https://blog.twitter.com/2014/what-fuels-a-tweets-engagement>

138. What are replies and mentions? [Internet]. *Twitter Help Center*. [cited 2015 Dec 8]. Available from: <https://support.twitter.com/articles/14023>

139. Replies and retweets on Twitter. *Sysomos* [Internet]. 2010 Sep; Available from: <https://www.sysomos.com/inside-twitter/twitter-retweet-stats>

140. Sakamoto T. An analysis on information diffusion by retweets in Twitter [Internet] [Thesis]. *Massachusetts Institute of Technology*; 2015 [cited 2015 Sep 3]. Available from: <http://dspace.mit.edu/handle/1721.1/97348>

141. Liu Y, Kliman-Silver C, Mislove A. The Tweets they are a-changin’: evolution of Twitter users and behavior. In: *Eighth International AAAI Conference on Weblogs and Social Media* [Internet]. 2014 [cited 2015 Sep 4]. p. 305–14. Figure 11, Figure 11: The percentage of tweets with different types of entities (bottom), and average number of entities for such tweets (top) over time. We observe increasing adoption of mentions, and an increasing likelihood of many hashtags per tweet; p. 312. Available from: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8043>

142. Liu Y, Kliman-Silver C, Mislove A. The Tweets they are a-changin’: evolution of Twitter users and behavior. In: *Eighth International AAAI Conference on Weblogs and*

- Social Media [Internet]. 2014 [cited 2015 Sep 4]. p. 305–14. Figure 8, The percentage of tweets of different types over time. Both RTs and Retweets represent retweets, with the former capturing manually created retweets by users. Native retweets were supported by Twitter starting in November 2009.; p. 310. Available from: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8043>
143. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer*. 2013 Oct 29;13:508.
144. Habel MA, Liddon N, Stryker JE. The HPV vaccine: a content analysis of online news stories. *J Womens Health* 2002. 2009 Mar;18(3):401–7.
145. Heavilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. *J Dent Res*. 2011 Sep 1;90(9):1047–51.
146. Schreier M. Qualitative content analysis. In: *The SAGE handbook of qualitative data analysis* [Internet]. London, United Kingdom: SAGE Publications Ltd; 2014 [cited 2015 Mar 13]. p. 170–83. Available from: <http://srmo.sagepub.com/view/the-sage-handbook-of-qualitative-data-analysis/n12.xml>
147. Williams C. Twitter cuts off data stream to British tech “unicorn” DataSift [Internet]. 2015 [cited 2015 Apr 21]. Available from: <http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/digital-media/11533275/Twitter-cuts-off-data-stream-to-British-tech-unicorn-DataSift.html>
148. Reporter DLT. Twitter partners with Datasift to unlock tweet archive [Internet]. BBC News. [cited 2015 Sep 19]. Available from: <http://www.bbc.co.uk/news/technology-17178022>
149. DataSift announces next generation historic [Internet]. Datasift Blog. [cited 2015 Apr 22]. Available from: <http://blog.datasift.com/2013/09/26/why-better-data-is-important-historic-access-across-social/>
150. Splunk docs. Splexicon. Index. [Internet]. [cited 2015 Aug 31]. Available from: <http://docs.splunk.com/Splexicon:Index>
151. Splunk Docs. Managing indexers and clusters of indexers. How indexing works. [Internet]. [cited 2015 Aug 25]. Available from: <http://docs.splunk.com/Documentation/Splunk/6.2.3/Indexer/Howindexingworks>
152. Splunk Docs. Splexicon. Event data. [Internet]. [cited 2015 Aug 25]. Available from: <http://docs.splunk.com/Splexicon:Eventdata>
153. Splunk docs. Search Manual. About search. [Internet]. [cited 2015 Aug 16]. Available from: <http://docs.splunk.com/Documentation/Splunk/6.2.4/Search/Aboutsearch>
154. Splunk Docs. Search Manual. About the Serach Processing Language. [Internet]. [cited 2015 Aug 16]. Available from: <http://docs.splunk.com/Documentation/Splunk/6.2.4/Search/Aboutthesearchlanguage>

155. Splunk Docs. Search Manual. About searching with time. [Internet]. [cited 2015 Aug 16]. Available from:
<http://docs.splunk.com/Documentation/Splunk/6.2.4/Search/Aboutsearchtimeranges>
156. Search manual. About transforming commands and searches. [Internet]. [cited 2015 Aug 16]. Available from:
<http://docs.splunk.com/Documentation/Splunk/6.2.4/Search/Aboutreportingcommands>
157. Splunk docs. Knowledge Manager Manual. About data models. [Internet]. [cited 2015 Aug 15]. Available from:
<http://docs.splunk.com/Documentation/Splunk/6.2.4/Knowledge/Aboutdatamodels>
158. Splunk Docs. Data model and Pivot tutorial. About Pivot. [Internet]. [cited 2015 Aug 16]. Available from:
<http://docs.splunk.com/Documentation/Splunk/6.2.4/PivotTutorial/Aboutpivotreports>
159. Splunk Docs. Splexicon. Event type. [Internet]. [cited 2015 Aug 31]. Available from:
<http://docs.splunk.com/Splexicon:Eventtype>
160. DataSift developers. Data Dictionary. [Internet]. [cited 2015 Feb 9]. Available from:
<http://dev.datasift.com/docs/resources/dictionary>
161. FAQs about Retweets (RT) [Internet]. Twitter Help Center. [cited 2015 Dec 8]. Available from: <https://support.twitter.com/articles/77606>
162. Centers for Disease Control and Prevention. Ebola Virus Disease. Previous Case Counts. [Internet]. [cited 2015 Aug 25]. Available from:
<http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/previous-case-counts.html>
163. World Health Organization. Ebola. Current Situation. Figure 1, Ebola cases in Guinea, Liberia, and Sierra Leone during the current Ebola Outbreak. [Internet]. Figure 1, Ebola cases in Guinea, Liberia, and Sierra Leone during the current Ebola Outbreak. [cited 2015 May 20]. Available from: <http://apps.who.int/ebola/>
164. WHO | Origins of the 2014 Ebola epidemic [Internet]. WHO. [cited 2015 Aug 25]. Available from: <http://www.who.int/csr/disease/ebola/one-year-report/virus-origin/en/>
165. WHO | WHO declares end of Ebola outbreak in Nigeria [Internet]. WHO. [cited 2015 Aug 25]. Available from:
<http://www.who.int/mediacentre/news/statements/2014/nigeria-ends-ebola/en/>
166. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *Am J Infect Control*. 2015 Jun;43(6):563–71.
167. Rodriguez-Morales AJ, Castañeda-Hernández DM, McGregor A. What makes people talk about Ebola on social media? A retrospective analysis of Twitter use. *Travel Med Infect Dis*. 2015 Jan;13(1):100–1.
168. About Twitter, Inc. | About [Internet]. Twitter About. [cited 2015 Mar 23]. Available from: <https://about.twitter.com/company>

169. Chew C, Eysenbach G. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. PLoS ONE. 2010 Nov 29;5(11):e14118.
170. Elena. 8 Funny Twitter Accounts to Follow in 2015 - FBlog [Internet]. [cited 2015 Sep 23]. Available from: <http://www.freemake.com/blog/funny-twitter-accounts-to-follow-in-2013/>
171. “Are you trying to be funny?” - Using Humour on Your Brand’s Social Presences | Social Media Marketing Blog - Social Marketing Blog [Internet]. Yomego. [cited 2015 Sep 23]. Available from: <http://www.yomego.com/blog/posts/2014/january/23/%25E2%2580%2598are-you-trying-to-be-funny%25E2%2580%2599-using-humour-on-your-brand%25E2%2580%2599s-social-presences/>
172. Whiting A, Williams D. Why people use social media: a uses and gratifications approach. Qual Mark Res Int J. 2013 Aug 30;16(4):366.
173. Magazine JKBN. Why do people tell sick jokes about tragedies? [Internet]. BBC News. [cited 2015 Sep 24]. Available from: <http://www.bbc.co.uk/news/magazine-12775389>
174. Alleviating Stress With Humour: a Literature Review [Internet]. Redorbit. [cited 2015 Sep 24]. Available from: http://www.redorbit.com/news/health/920164/alleviating_stress_with_humour_a_literature_review/
175. Kuiper NA, Martin RA, Joan L. Coping humour, stress, and cognitive appraisals. Can J Behav Sci Can Sci Comport. 1993;25(1):81–96.
176. Why We Laugh [Internet]. Psychology Today. [cited 2015 Sep 24]. Available from: <http://www.psychologytoday.com/blog/happiness-in-world/201101/why-we-laugh>
177. WHO | Goodwill ambassadors [Internet]. WHO. [cited 2015 Sep 25]. Available from: http://www.who.int/goodwill_ambassadors/en/
178. UN Women Goodwill Ambassador Emma Watson [Internet]. headQuarters. [cited 2015 Sep 25]. Available from: <http://www.unwomen.org/en/partnerships/goodwill-ambassadors/emma-watson>
179. Goodwill Ambassadors & Advocates | UNICEF People [Internet]. UNICEF. [cited 2015 Sep 25]. Available from: http://www.unicef.org/people/people_ambassadors.html
180. Nepal earthquake: Shah Rukh Khan shares emergency contact numbers [Internet]. Zee News. [cited 2015 Sep 25]. Available from: http://zeenews.india.com/entertainment/celebrity/nepal-earthquake-shah-rukh-khan-shares-emergency-contact-numbers_1586464.html
181. Sedhai S, Sun A. HSpam14: A Collection of 14 Million Tweets for Hashtag-Oriented Spam Research. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval [Internet]. New York, NY, USA: ACM; 2015 [cited 2015 Sep 14]. p. 223–32. Available from: <http://doi.acm.org/10.1145/2766462.2767701>

182. Mislove A, Jørgensen SL, Ahn Y-Y, Onnela J-P, Rosenquist JN, Mislove A, et al. Understanding the Demographics of Twitter Users. Proc Fifth Int AAAI Conf Weblogs Soc Media. 2011;554–7.
183. Twitter: number of monthly active users 2014 | Statistic [Internet]. Statista. [cited 2015 Mar 22]. Available from: <http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>
184. Twitter glossary of terms [Internet]. Media. [cited 2015 Mar 29]. Available from: <https://media.twitter.com/best-practice/glossary>
185. Twitter Help Center. The Twitter Rules [Internet]. [cited 2015 Mar 30]. Available from: <https://support.twitter.com/articles/18311-the-twitter-rules>
186. DataSift.Sources. [Internet]. [cited 2015 Aug 25]. Available from: <http://app.datasift.com/source>

Appendices

Appendix 1. Twitter glossary of terms.

Appendix 2. Data augmentations provided by DataSift.

Appendix 3. Example of a DataSift historic preview output used in the study (covering the period 1 June 2014 – 1 July 2014).

Appendix 4. Example of the full event data output in Splunk interface.

Appendix 5. Statement of approval of the Research and Ethics Committee of the Oxford Internet Institute.

Appendix 6. Features of the top 20 retweeted accounts over the whole study period and each of the sub-periods used for their categorisation.

Appendix 7. Most retweeted tweets and the types of accounts that posted them in each of the study sub-periods.

Appendix 8. Features of the top 20 mentioned accounts over the whole study period and each of the sub-periods used for their categorisation.

Appendix 9. Features of the top 20 accounts that received the highest number of replies over the whole study period and each of the sub-periods used for their categorisation.

Appendix 1. Twitter glossary of terms. (184)

Term	Meaning
Alerts	Twitter Alerts enable public safety agencies to inform people during emergencies by highlighting critical time-sensitive content with notifications and a unique look.
Bio	Bio is a short (up to 160 characters) personal description that appears in user's profile that serves to characterize user's persona on Twitter.
Block	If a user blocks a Twitter user, that account will be unable to follow the user or add the user to their Twitter lists, and the user will not receive a notification if they mention the user in a tweet.
Cashtag	A cashtag is a company ticker symbol preceded by the US dollar sign, e.g. \$TWTR. When a user click on a cashtag, a user will see other tweets mentioning that same ticker symbol.
Collection	A group of tweets a user puts together and shares with their followers. Each collection is public, embeddable and can be curated in real time.
Deactivation	If a user deactivates their account, it goes into a queue for permanent deletion from Twitter in 30 days. The user may reactivate their account within the 30-day grace period.
Discover	This feature surfaces personalized content tailored to user's interests.
Favourite	Taping the star icon to favourite a tweet and the author demonstrates that a user liked it. Favouring a tweet indicates that the user liked a specific tweet. The user can find all of their favourite Tweets by clicking on the favourites link on their profile page.
Follow	A follow is the result of someone following Twitter user account. The user can see how many follows (or followers) they have from their Twitter profile.
Following	Subscribing to a Twitter account is called "following". To start following, a user must click the Follow button next to the user name or on their profile page to see their Tweets as soon as they post something new. Anyone on Twitter can follow or unfollow anyone else at any time, with the exception of blocked accounts.
Follow count	This count reflects how many people a user follow and how many follow them; these numbers are found on Twitter user profile.
Follower	A follower is another Twitter user who has followed a user to receive their Tweets in their Home stream.
Geo-location, geo-tagging:	Adding a location to user's tweet (a geo-location or geo-tag) tells those who see user's tweet where they were when they posted that Tweet.
Hashtag	A hashtag is any word or phrase immediately preceded by the # symbol. When a user clicks on a hashtag, they will see other Tweets containing the same keyword or topic.
Home	Home is user's real-time stream of tweets from those they follow.
Lists	From user's account, they can create a group list of other Twitter users by the topic or interest (e.g., a list of friends, co-workers, celebrities, and athletes). Twitter lists also contain a timeline of tweets from the specific users that were added to the list, offering the user a way to follow individual accounts as a group on Twitter.
Mention	Mentioning other users in a tweet by including the @ sign followed directly by their username is called a "mention". Also refers to tweets in which the user's @username was included.
Tweets (messages)	Messages have a 140-character limit and can contain text, hashtags, links, photos and video.
MT	Similar to RT (retweet), an abbreviation for "Modified Tweet". Placed before the Retweeted text when users manually retweet a message with modifications, for example shortening a tweet.
Parody	A user can create parody accounts on Twitter to spoof or make fun of something in

	jest, as well as commentary and fan accounts. These accounts must disclose that they are parody, fan or commentary accounts in order to comply with Twitter policy against impersonation.
Profile	User profile displays information they choose to share publicly, as well as all of the tweets they have posted. Their profile along with their @username identifies them on Twitter.
Profile photo	User personal image found under the “Me” icon. It is also the picture that appears next to each of the user tweets.
Promoted tweets	Promoted tweets are tweets that are paid for by Twitter advertisers. These appear in user Home timeline, at the top of search results on Twitter and elsewhere on the platform, and are clearly marked as “Promoted.”
Pinned tweets	A user can pin a tweet to the top of their profile page, to keep something important to them above the flow of time-ordered tweets.
Reply	A response to another user’s tweet that begins with the @username of the person a user is replying to is known as a reply. It is done by clicking the “reply” button next to the tweet user would like to respond to.
Retweet	A tweet that a user forwards to their followers is known as a retweet. Often used to pass along news or other valuable discoveries on Twitter, retweets always retain original attribution.
Retweeting	The act of sharing another user’s Tweet to all of user followers by clicking on the retweet button.
Spam	Refers to a variety of prohibited behaviours that violate the Twitter Rules. (185) Spam can be generally described as unsolicited, repeated actions that negatively impact other users.
Search	A box on Twitter user homepage that allows them to search all public tweets for keywords, usernames, hashtags, cashtags or subjects. Searches can also be performed at search.twitter.com.
Suspended	Suspended accounts have been prohibited from using Twitter, generally for breaking Twitter Terms of Service.
Trends	A Trend is a topic or hashtag determined algorithmically to be one of the most popular on Twitter at that moment. A user can choose to tailor Trends based on their location and whom they follow.
Tweet button	Anyone can add a tweet button to their website. Clicking this button lets Twitter users post a tweet with a link to that site.
Username	A username is how you’re identified on Twitter, and is always preceded immediately by the @ symbol.
URL	A URL is a web address that points to a unique page on the internet.
Unfollow	Stopping following someone.
Verification	A process whereby a Twitter account receives a blue check icon to indicate that the creator of these Tweets is a legitimate source. Verified users include public figures and those who may have experienced identity confusion on Twitter.
Who to follow	Who to follow is an automated list of recommended accounts Twitter thinks a user might find interesting, based on the types of accounts they already follow and who those people follow.

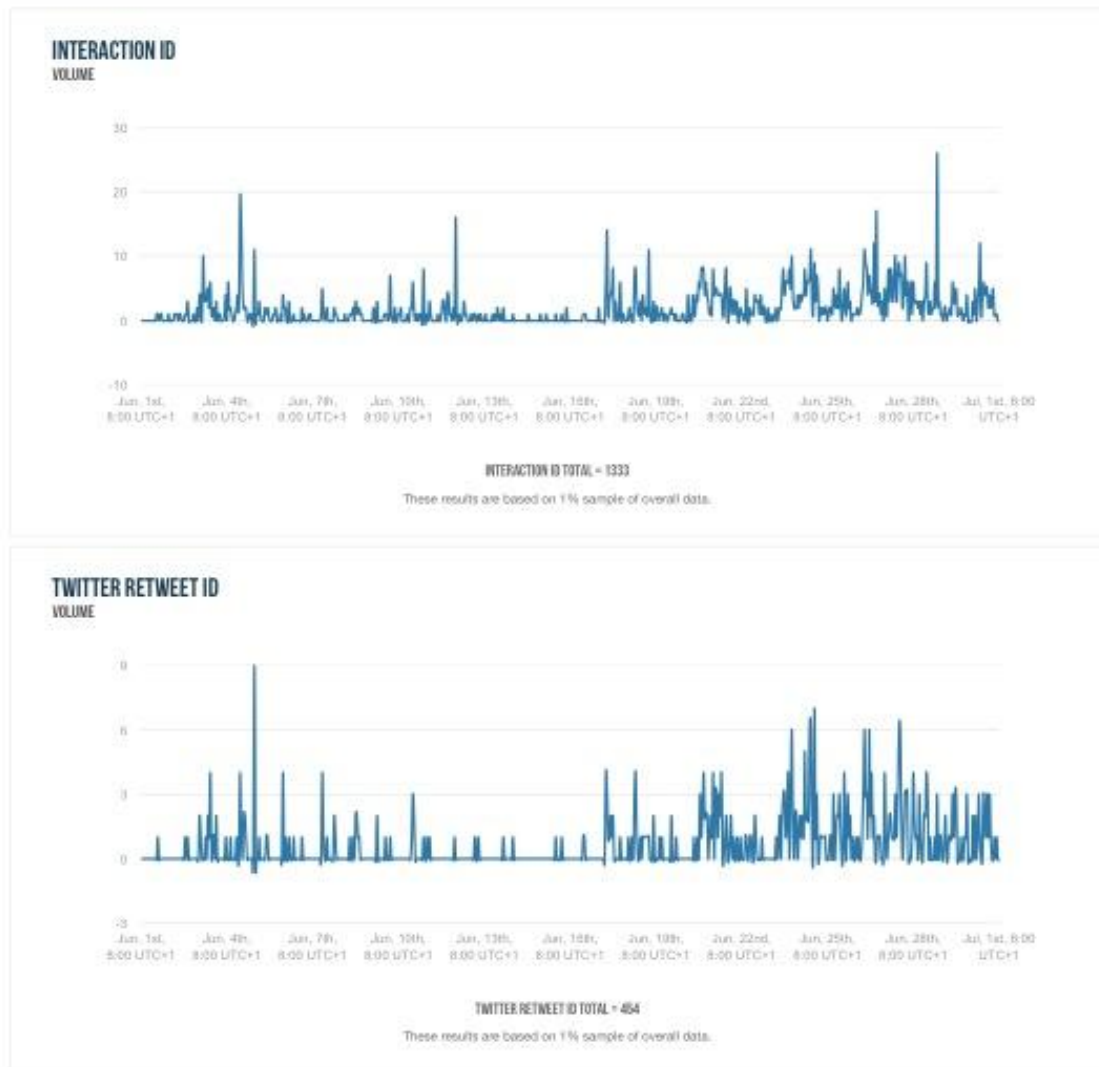
Appendix 2. Data augmentations provided by DataSift. (186)

Augmentation type	Description
Interaction augmentation	The interaction augmentation groups common features (main text, the creation time, the author, the links and other) from different data feeds together in order to make it easier to write generic queries across all the feeds during the analysis.
Gender augmentation	The gender augmentation determines author's gender by analysing names and home countries with the use of trigram text analysis and a database covering over 55 languages and geographical regions. The outputs are: female, male, unisex, mostly male, and mostly female.
Klout score	Klout measures an author's online influence by evaluating the data from different social networks in order to calculate an overall measure of influence on a scale from 1 to 100. The Klout Score uses data from a variety of social networks to calculate an overall measure of influence on a scale from 1 to 100.
Language augmentation	Language augmentation automatically detects the language in which the post is written and supports more than 140 languages. Moreover, it provides the confidence of the estimated accuracy level of the detection.
Links augmentation	Links augmentation looks at any links within the content and resolves them to their final endpoint. It also aggregates the link content so that the filtering can be performed against the content of the page that the link was pointing at.
Saliency topics and entities	The Saliency Topics augmentation lists the topics it finds in a post. Saliency Topics are derived from an analysis of all the content in Wikipedia. The engine independently analyses the content and the title. The Saliency Entities data source lists the entities it finds in a post. Entities are typically proper nouns (people, companies, products, places) but can also include special patterns like URLs, phone numbers, addresses, currency amounts, and quotes.
Sentiment	Sentiment augmentation delivers a measure of the positive or negative sentiment in a post. Within DataSift, sentiment is typically scored from -20 to 20. A score of zero is neutral, scores below zero indicate negative sentiment, and scores above zero indicate positive sentiment. For example, a score of -1 is mildly negative while a score of 15 is strongly positive. Sentiment works with posts written in English, Spanish, French, and Portuguese.
Trends	DataSift receives trend data directly from Twitter (which topics are trending by country and city) so trends are restricted to streams that use data from Twitter.

Appendix 3. Example of a DataSift historic preview output used in the study (covering the period 1 June 2014 – 1 July 2014).

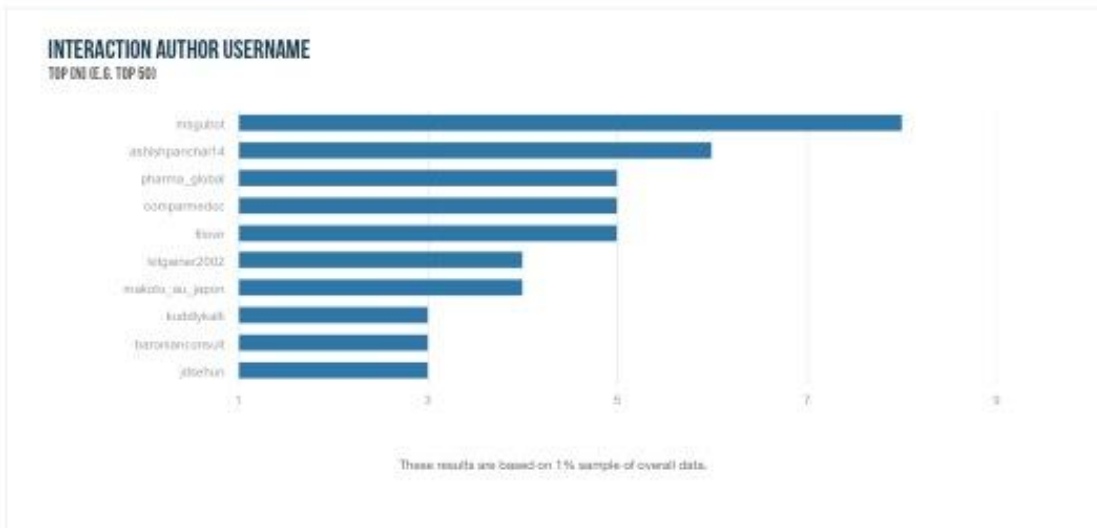
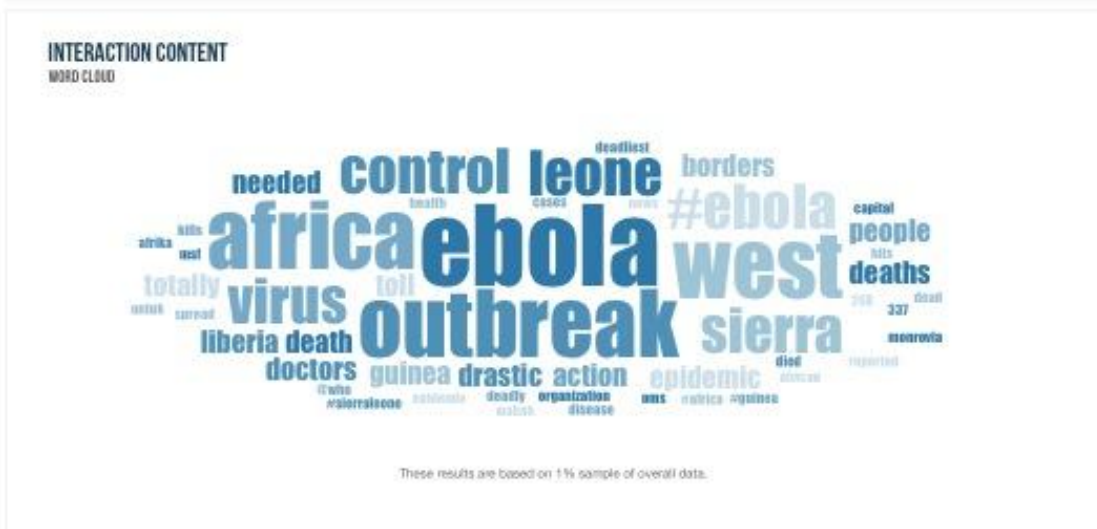
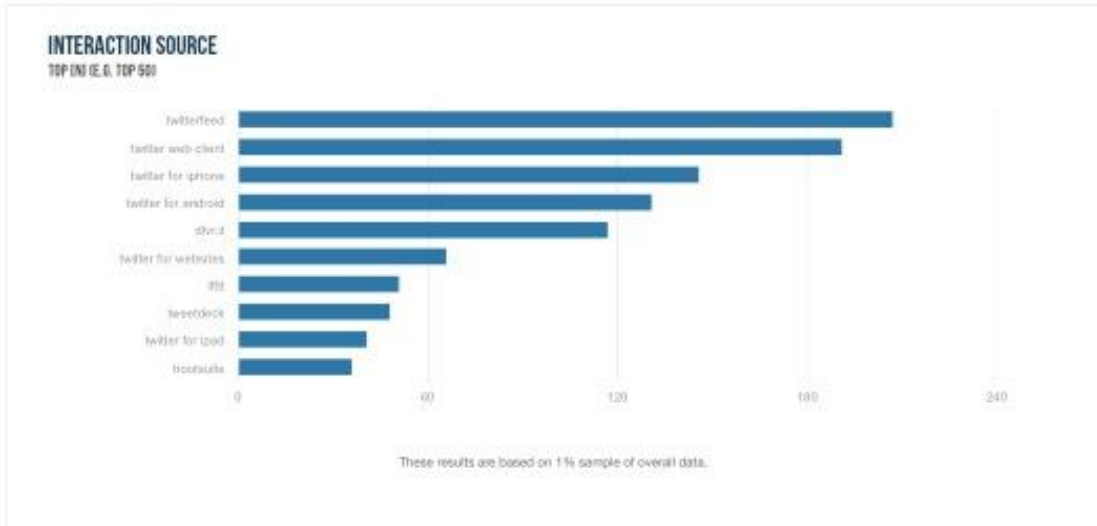
20.04.2015

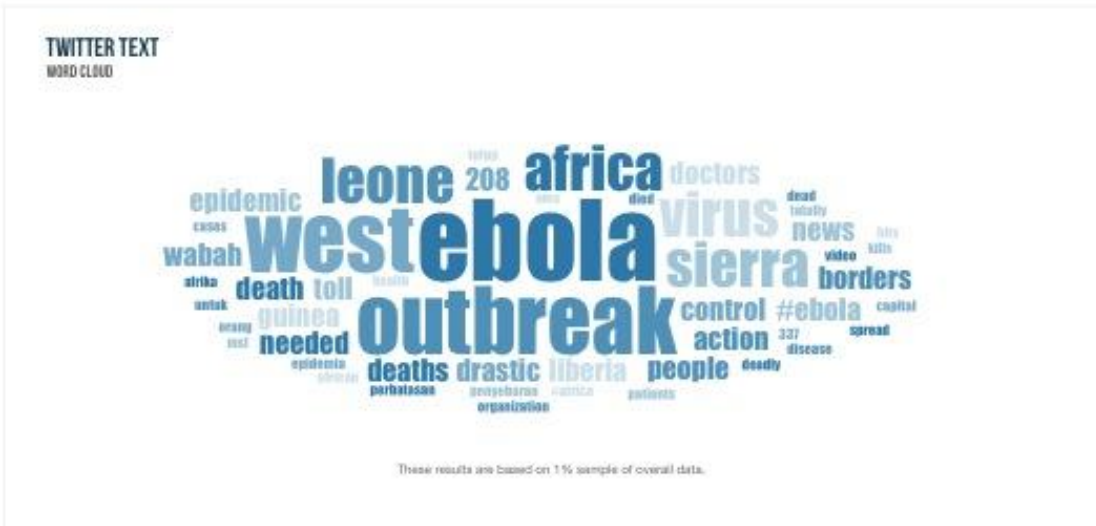
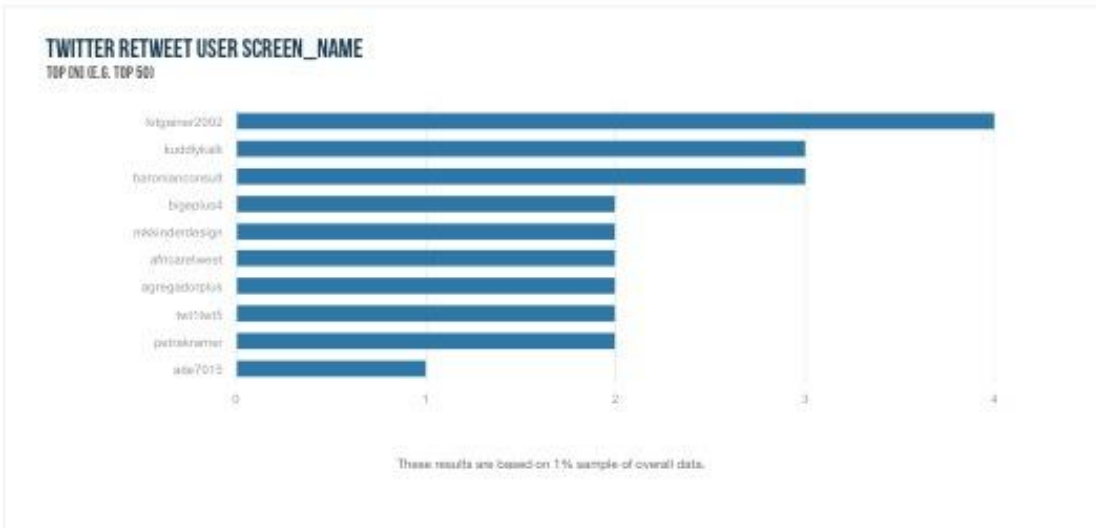
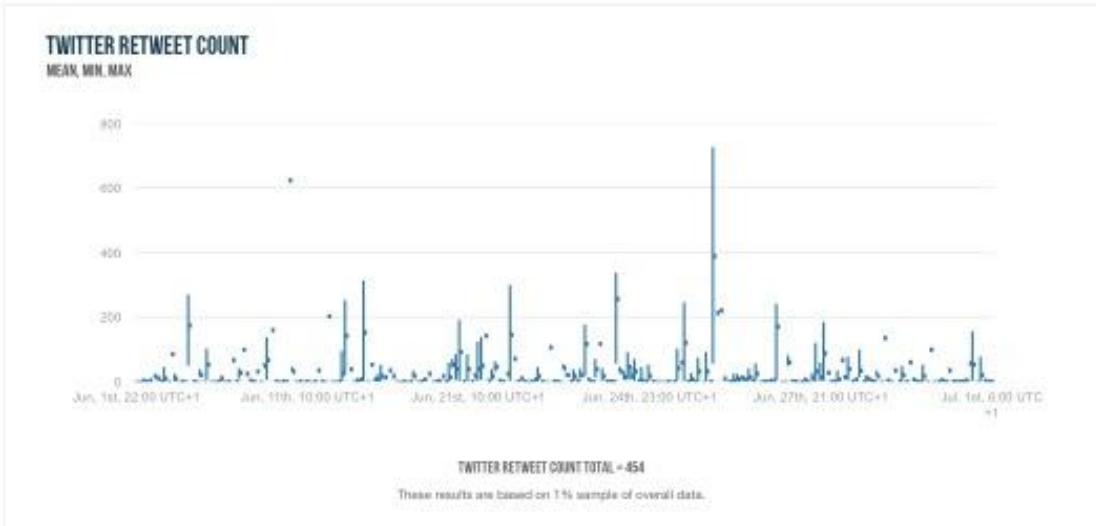
Use 7 June - Historic Preview - DataSift

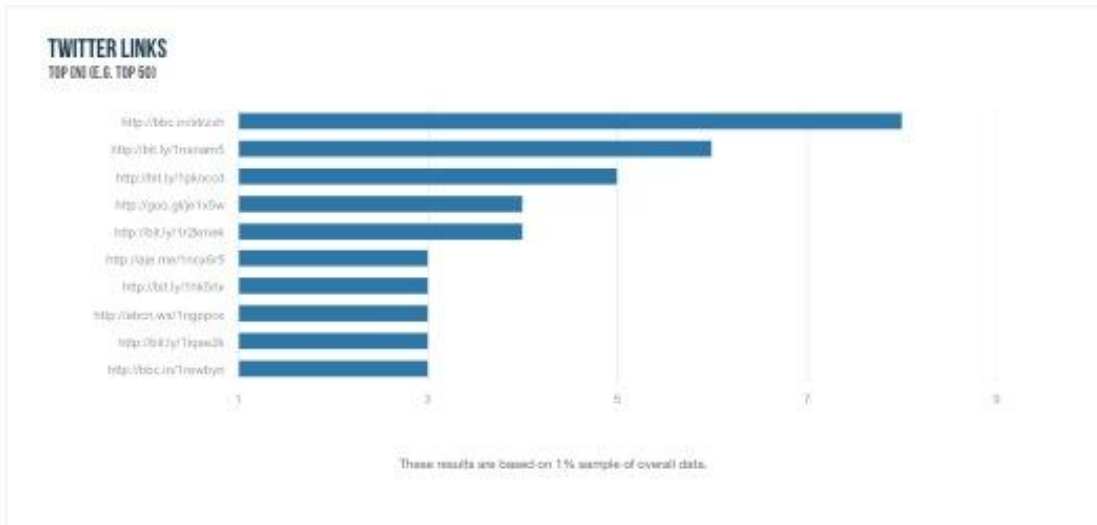
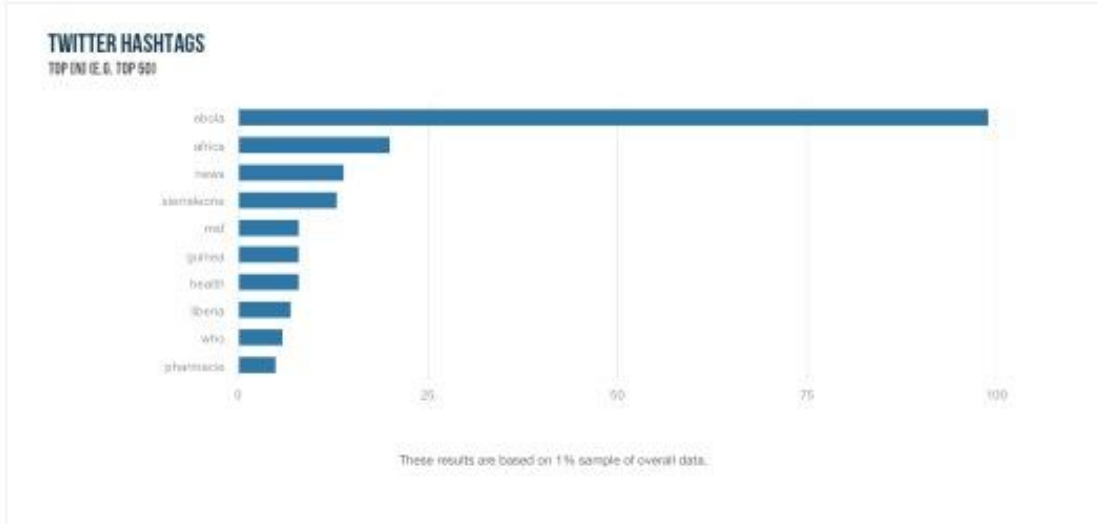


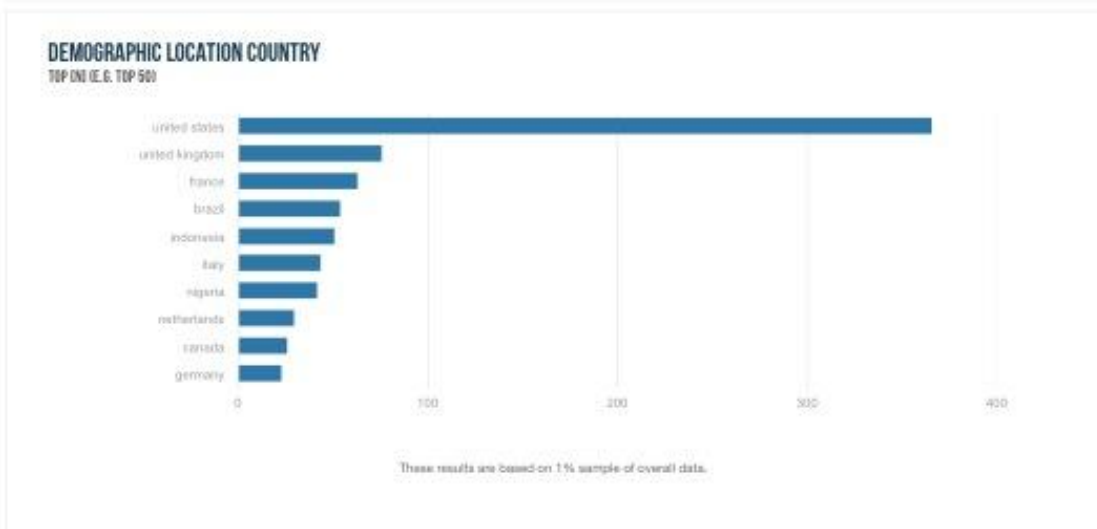
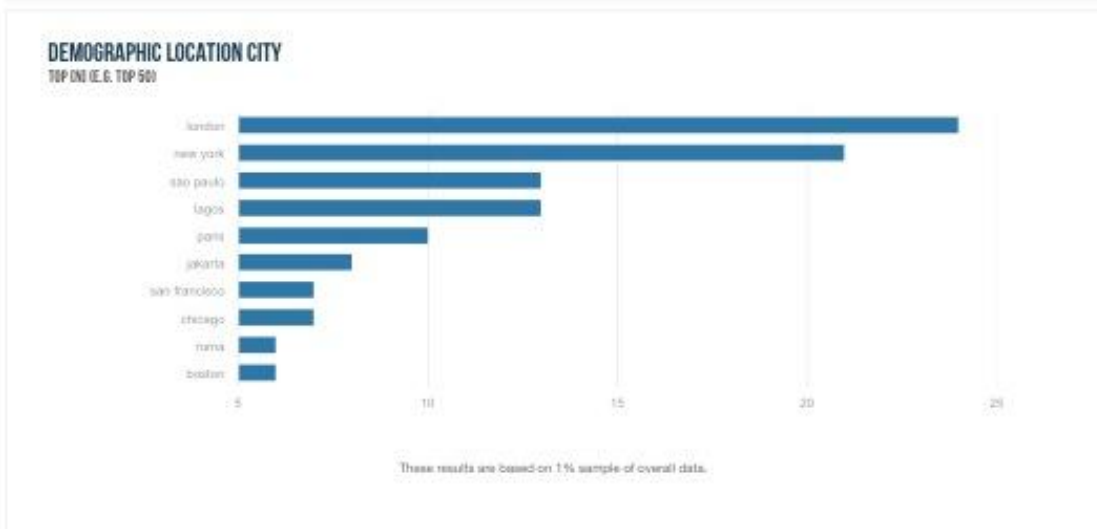
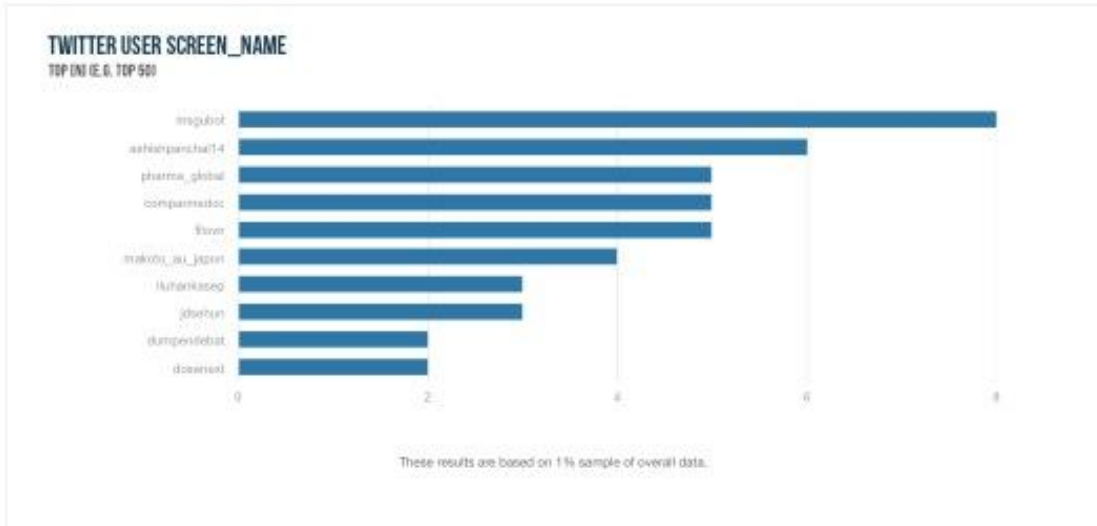
<https://app.datasift.com/streams/158641/use/preview/historic>

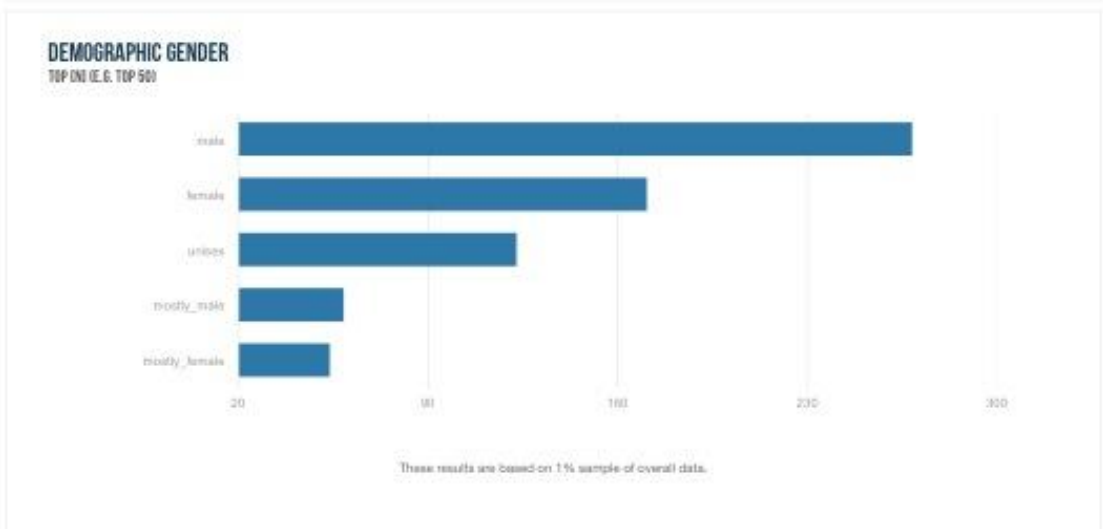
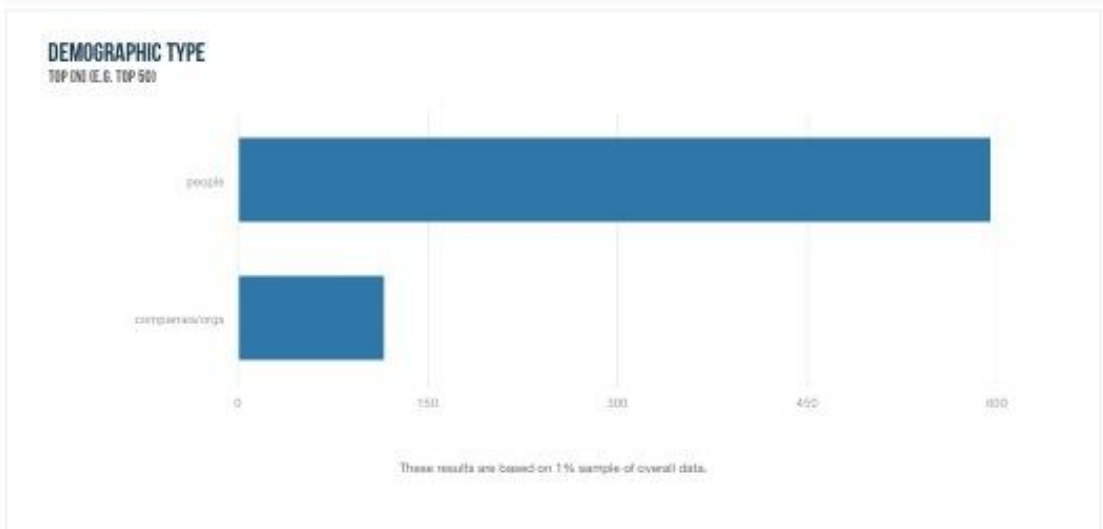
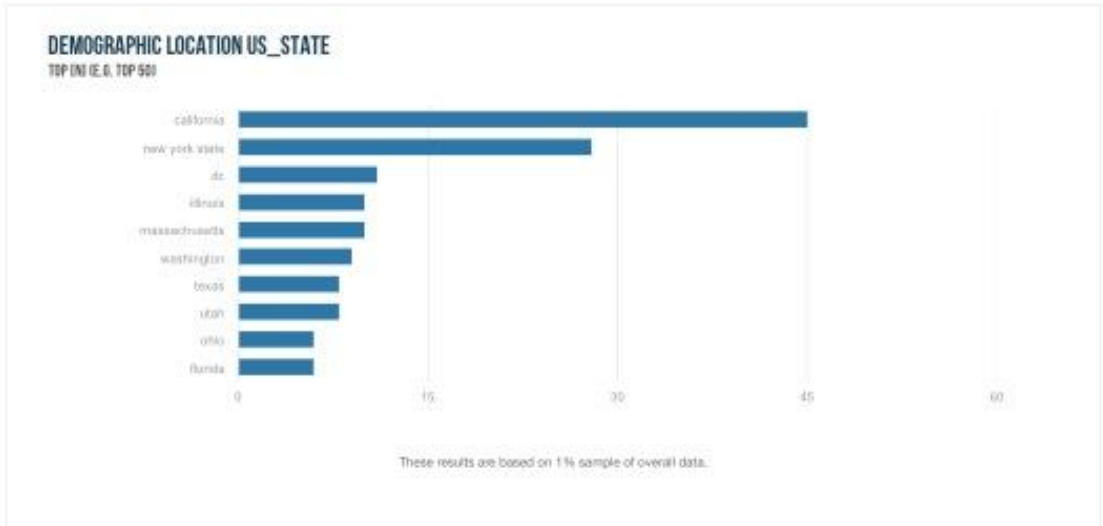
1/9

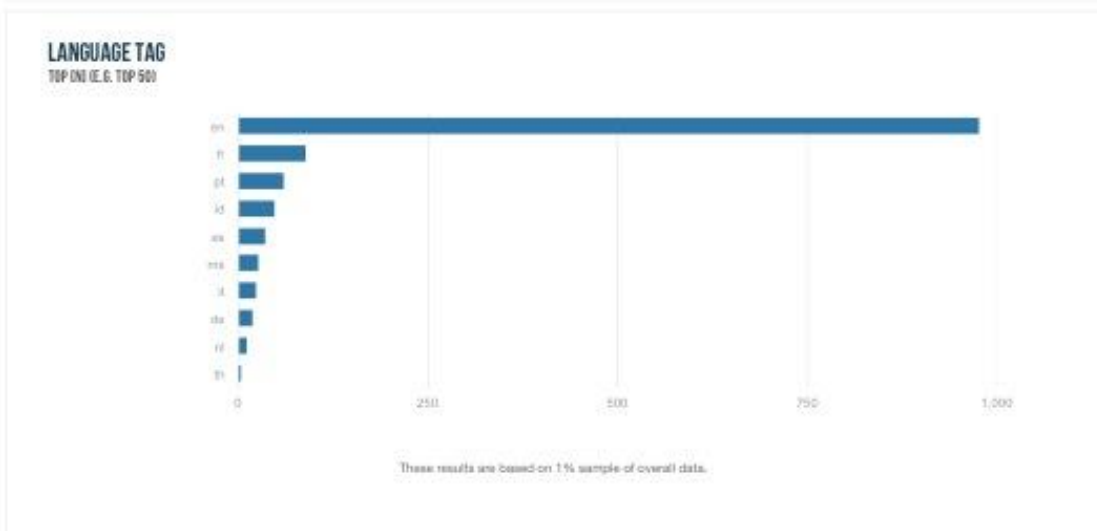
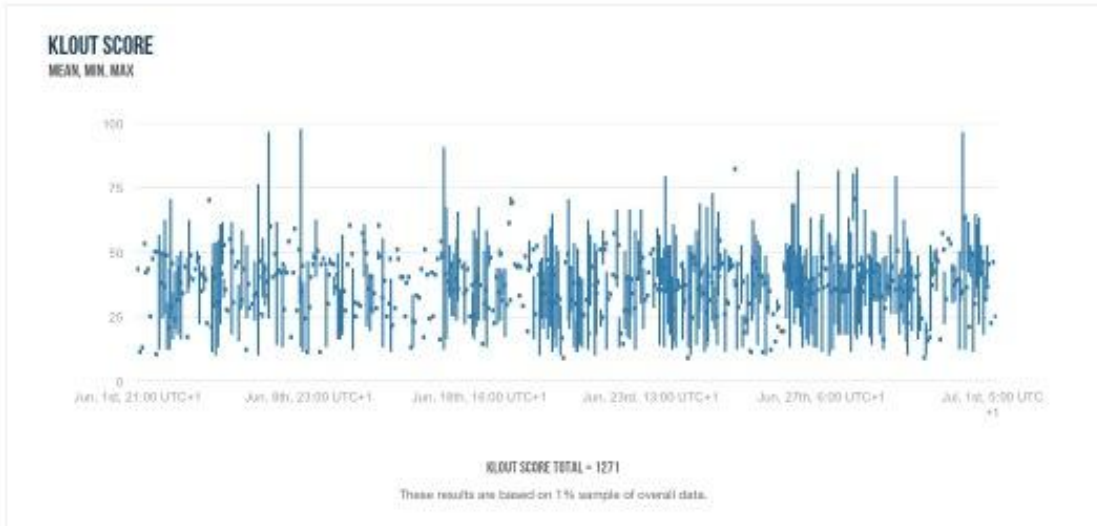


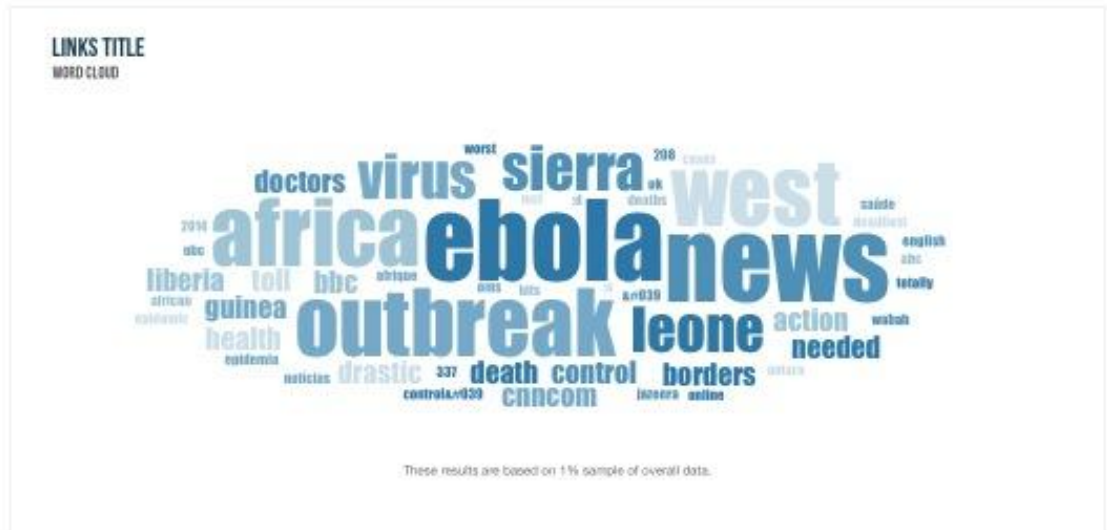












Appendix 4. Example of the full event data output in Splunk interface.

5/12/1 5 6:17:08 .000 AM	<pre>{ [-] interaction: { [-] author: { [-] avatar: http://pbs.twimg.com/profile_images/37880000419340256/718d9de1b7 a25577c6520f3e1073ce88_normal.jpeg id: 141152082 language: en link: https://twitter.com/JME_BMJ name: Medical Ethics username: JME_BMJ } content: RT @WHO: WHO declares the end of the #Ebola outbreak in #Liberia Liberia #StaysVigilant http://t.co/WHPr2dT3xJ created_at: Tue, 12 May 2015 06:17:08 +0000 hashtags: [[-] Ebola Liberia StaysVigilant] id: 1e4f86e84704aa00e074d2f775255d6c link: https://twitter.com/JME_BMJ/status/598008989100584960 mention_ids: [[-] 14499829] mentions: [[-] WHO] received_at: 1431411428.387 schema: { [-] version: 3 } source: Twitter Web Client type: twitter } klout: { [-] score: 52 } language: { [-] confidence: 98 tag: en tag_extended: en } twitter: { [-] id: 598008989100584960 retweet: { [-] count: 3046 created_at: Tue, 12 May 2015 06:17:08 +0000 filter_level: low hashtags: [[-]</pre>
--------------------------------------	--

```

Ebola
Liberia
StaysVigilant
]
id: 598008989100584960
lang: en
media: [ [-]
  { [-]
    display_url: pic.twitter.com/WHPr2dT3xJ
    expanded_url: http://twitter.com/WHO/status/596977292007624704/photo/
1
    id: 596977278199824400
    id_str: 596977278199824384
    media_url: http://pbs.twimg.com/media/CEjjq-4UgAA2kyB.jpg
    media_url_https: https://pbs.twimg.com/media/CEjjq-4UgAA2kyB.jpg
    sizes: { [+]}
    type: photo
    url: http://t.co/WHPr2dT3xJ
  }
  { [-]
    display_url: pic.twitter.com/WHPr2dT3xJ
    expanded_url: http://twitter.com/WHO/status/596977292007624704/photo/
1
    id: 596977278199824400
    id_str: 596977278199824384
    media_url: http://pbs.twimg.com/media/CEjjq-4UgAA2kyB.jpg
    media_url_https: https://pbs.twimg.com/media/CEjjq-4UgAA2kyB.jpg
    sizes: { [+]}
    type: photo
    url: http://t.co/WHPr2dT3xJ
  }
]
mention_ids: [ [-]
  14499829
]
mentions: [ [-]
  WHO
]
source: <a href="http://twitter.com" rel="nofollow">Twitter Web
Client</a>
text: WHO declares the end of the #Ebola outbreak in #Liberia Liberia
#StaysVigilant http://t.co/WHPr2dT3xJ
user: { [-]
  created_at: Fri, 07 May 2010 09:16:34 +0000
  description: The #1 ranked international #bioethics journal by Google
Scholar Metrics. Impact Factor = 1.691. H5-index = 29. Tweets & RTs are not
endorsements.
  favourites_count: 31
  followers_count: 7461

```

```
    friends_count: 2679
    geo_enabled: false
    id: 141152082
    id_str: 141152082
    lang: en
    listed_count: 263
    name: Medical Ethics
    profile_image_url: http://pbs.twimg.com/profile_images/378800004193402
56/718d9de1b7a25577c6520f3e1073ce88_normal.jpeg
    profile_image_url_https: https://pbs.twimg.com/profile_images/378800000
419340256/718d9de1b7a25577c6520f3e1073ce88_normal.jpeg
    screen_name: JME_BMJ
    statuses_count: 1880
    time_zone: Amsterdam
    url: http://jme.bmj.com
    utc_offset: 7200
    verified: false
  }
}
retweeted: { [-]
  created_at: Sat, 09 May 2015 09:57:32 +0000
  id: 596977292007624704
  source: <a href="http://twitter.com/download/iphone"
rel="nofollow">Twitter for iPhone</a>
  user: { [-]
    created_at: Wed, 23 Apr 2008 19:56:27 +0000
    description: Official Twitter account of the World Health Organization,
the United Nations' health agency
    favourites_count: 3630
    followers_count: 2356697
    friends_count: 964
    geo_enabled: true
    id: 14499829
    id_str: 14499829
    lang: en
    listed_count: 19122
    location: Geneva, Switzerland
    name: WHO
    profile_image_url: http://pbs.twimg.com/profile_images/2189537674/WHOLo
go_normal.png
    profile_image_url_https: https://pbs.twimg.com/profile_images/218953767
4/WHOLogo_normal.png
    screen_name: WHO
    statuses_count: 15309
    time_zone: Bern
    url: http://www.who.int
    utc_offset: 7200
    verified: true
  }
}
}
```

Appendix 5. Statement of approval of the Research and Ethics Committee at the Oxford Internet Institute.



Oxford Internet Institute
University of Oxford
1 St Giles'
Oxford
OX1 3JS

14 August 2015

Re: CUREC approval for '#Ebola outbreak on Twitter: lessons learnt'

Dear Tatiana,

I am pleased to write that the Research & Ethics Committee at the Oxford Internet Institute have approved your submission of the CUREC 1A form for your proposed project '#Ebola outbreak on Twitter: lessons learnt', approval being granted on 10 August 2015.

Your CUREC approval reference is: OII C1A 15-045.

On behalf of the Research & Ethics Committee, I wish you every success with your research.

Best wishes,

Dr Tim Davies
Research Facilitator
Oxford Internet Institute
University of Oxford

+44 (0)1865 612334
tim.davies@oii.ox.ac.uk

oxford internet institute
university of oxford
one st giles oxford ox1 3js
united kingdom
tel +44(0)1865 287210 fax +44(0)1865 287211
enquiries@oii.ox.ac.uk
www.oii.ox.ac.uk



Appendix 6. Features of the top 20 retweeted accounts over the whole study period and each of the sub-periods used for their categorisation.

Table 1. Categorisation of the top 20 retweeted accounts in the full study period.

Twitter user screen name	Number of retweets	Prevalence of retweet number among the top 20 retweeted accounts	Verification by Twitter	Highest number of retweets within the period	Average number of retweets within the period	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WorldStarFunny	215,293	9.4%	no	23,811	11,446	865,369	29	Humour
nytimes	190,536	8.3%	yes	3,394	1,663	11,369,883	1,338	News
WHO	177,483	7.7%	yes	3,195	1,586	1,173,156	1,781	(Inter)governmental organisation/NGO
cnnbrk	152,049	6.6%	yes	5,294	2,623	16,128,604	260	News
ItsYaVirusEBOLA	150,291	6.5%	no	28,139	13,527	357	77	Humour
Dory	149,657	6.5%	no	23,553	11,452	669,753	39	Humour
UNICEF	120,562	5.2%	yes	3,293	1,612	2,595,866	780	(Inter)governmental organisation/NGO
FunnyPicsDepot	119,671	5.2%	no	10,288	4,99	1,178,130	27	Humour
CNN	99,59	4.3%	yes	25,191	12,058	12,264,670	157	News
BBCBreaking	95,452	4.2%	yes	3,827	1,902	9,222,667	165	News
AP	94,811	4.1%	yes	3,408	1,665	3,096,719	672	News
A3Noticias	92,375	4.0%	yes	11,964	5,494	625,058	237	News
YaBoyEbola	90,757	4.0%	no	13227	5,976	0	44	Humour
TIME	84,996	3.7%	yes	11,207	5,453	5,572,684	743	News
BBCWorld	84,296	3.7%	yes	7,684	3,734	5,901,375	396	News
LeadershipNGA	79,086	3.4%	no	218	110	79,945	21,782	Automated tweet aggregator
CloydRivers	78,196	3.4%	no	8,08	3,947	745,546	32	Twitter@celebrity
ABC	75,229	3.3%	yes	9,785	4,734	3,096,995	573	News
RelatableQuote	74,875	3.3%	no	18,034	8,809	2,779,084	7	Humour
CauseWereGuys	74,137	3.2%	no	5,518	2,703	1,656,433	31	Twitter@celebrity

Table 2. Categorisation of the top 20 retweeted accounts in the 1st sub-period.

Twitter user screen name	Number of retweets	Prevalence of retweets number among the top 20 retweeted accounts	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
FluTrackers	2,156	33.3%	no	3,249	6,48	Infectious disease tracking
RonanKelly13	1,071	16.5%	no	374	2,527	Infectious disease tracking
ironorehopper	751	11.6%	no	333	6,638	Automated tweet aggregator
GertvanderHoek	654	10.1%	no	659	1,332	Infectious disease tracking
ebolasound	291	4.5%	no	2,835	753	Twitter celebrity
KRFront	221	3.4%	no	4,428	1,162	Twitter celebrity
gengelis	206	3.2%	no	2,359	1	Twitter celebrity
Dr_Ebola	118	1.8%	no	0	1	Other non-verified
HuffingtonPost	113	1.7%	yes	3,842,480	697	News
healthgovau	105	1.6%	yes	13,112	52	(Inter)governmental organisation/NGO
MaraWritesStuff	98	1.5%	yes	74,107	18	Celebrity
BatmanOfNight	95	1.5%	no	18,505	18	Twitter celebrity
Vice_Is_Hip	95	1.5%	no	60,872	37	Humour
NGIndonesia	87	1.3%	no	128,287	401	Twitter celebrity
medschooladvice	80	1.2%	no	61,212	3	Twitter celebrity
yotobi	77	1.2%	no	46,965	1	Twitter celebrity
maxkeiser	73	1.1%	no	94,313	15	Twitter celebrity
SeXapellaHatun	69	1.1%	no	2,611	1	Twitter celebrity
YourAnonNews	64	1.0%	no	1,200,222	35	Twitter celebrity
HuffPostGreen	59	0.9%	yes	221,153	124	News

Table 3. Categorisation of the top 20 retweeted accounts in the 2nd sub-period.

Twitter user screen name	Number of retweets	Prevalence of retweets number among the top 20 retweeted accounts	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WHO	20,667	34.3%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
BBCWorld	5,499	9.1%	yes	5,901,375	396	News
cnni	3,886	6.5%	yes	1,895,824	186	News
BBCAfrica	3,516	5.8%	yes	369,999	1,219	News
FluTrackers	3,185	5.3%	no	3,249	6,48	Infectious disease tracking
RT_com	2,058	3.4%	yes	630,043	492	News
AP	1,827	3.0%	yes	3,096,719	672	News
AJEnglish	1,751	2.9%	yes	1,904,039	232	News
AFP	1,684	2.8%	yes	169,25	523	News
CNN	1,666	2.8%	yes	12,264,670	157	News
BBCBreaking	1,602	2.7%	yes	9,222,667	165	News
Reuters	1,556	2.6%	yes	4,097,821	314	News
afpfr	1,555	2.6%	yes	339,782	811	News
UN	1,531	2.5%	yes	2,475,978	571	(Inter)governmental organisation/NGO
DRUDGE_REPORT	1,515	2.5%	no	647,372	926	Automated tweet aggregator
NigeriaNewsdesk	1,450	2.4%	yes	520,795	6,719	News
RonanKelly13	1,348	2.2%	no	374	2,527	Infectious disease tracking
makoto_au_japon	1,339	2.2%	no	1,388	8,228	Twitter celebrity
cnnbrk	1,283	2.1%	yes	16,128,604	260	News
Crof	1,273	2.1%	no	2,073	4,602	Twitter celebrity

Table 4. Categorisation of the top 20 retweeted accounts in the 3rd sub-period.

Twitter user screen name	Number of retweets	Prevalence of retweets number among the top 20 retweeted accounts	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WHO	86,445	13.9%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
NigeriaNewsdesk	59,937	9.6%	yes	520,795	6,719	News
nytimes	56,685	9.1%	yes	11,369,883	1,338	News
cnnbrk	39,608	6.4%	yes	16,128,604	260	News
emmkaff	34,456	5.5%	no	2,175	12	Twitter celebrity
BBCBreaking	33,985	5.5%	yes	9,222,667	165	News
BBCAfrica	32,278	5.2%	yes	369,999	1,219	News
BallinInHD	29,427	4.7%	no	45,098	26	Twitter celebrity
AP	29,113	4.7%	yes	3,096,719	672	News
Reuters	28,221	4.5%	yes	4,097,821	314	News
BBCWorld	27,296	4.4%	yes	5,901,375	396	News
NTANewsNow	25,497	4.1%	no	41,669	2,243	Twitter celebrity
washingtonpost	20,415	3.3%	yes	2,482,703	609	News
UNICEF	19,536	3.1%	yes	2,595,866	780	(Inter)governmental organisation/NGO
ABC	19,504	3.1%	yes	3,096,995	573	News
CNN	17,672	2.8%	yes	12,264,670	157	News
TIME	17,663	2.8%	yes	5,572,684	743	News
WSJ	15,656	2.5%	yes	4,753,218	420	News
NBCNews	15,495	2.5%	yes	1,397,688	487	News
TheTweetOfGod	14,912	2.4%	no	1,480,204	9	Humour

Table 5. Categorisation of the top 20 retweeted accounts in the 4th sub-period.

Twitter user screen name	Number of retweets	Prevalence of retweets number among the top 20 retweeted accounts	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WorldStarFunny	198,719	12.3%	no	865,369	29	Humour
ItsYaVirusEBOLA	150,291	9.3%	no	357	77	Humour
Dory	137,875	8.5%	no	669,753	39	Humour
FunnyPicsDepot	108,752	6.7%	no	1,178,130	27	Humour
cnnbrk	97,287	6.0%	yes	16,128,604	260	News
nytimes	93,484	5.8%	yes	11,369,883	1,338	News
YaBoyEbola	87,928	5.4%	no	0	44	Humour
A3Noticias	79,897	5.0%	yes	625,058	237	News
CauseWereGuys	72,192	4.5%	no	1,656,433	31	Twitter celebrity
CNN	69,862	4.3%	yes	12,264,670	157	News
CloydRivers	66,539	4.1%	no	745,546	32	Twitter celebrity
YABOYLILB	57,300	3.6%	no	353,473	23	Humour
_Kyheim	52,471	3.3%	no	2,406	5	Twitter celebrity
WORIDSTARHIPHOP	51,795	3.2%	no	242,816	38	Humour
RelatableQuote	50,450	3.1%	no	2,779,084	7	Humour
girlposts	50,017	3.1%	no	4,732,758	5	Twitter celebrity
WeLoveRobDyrdek	49,228	3.1%	no	463,861	34	Humour
AP	47,793	3.0%	yes	3,096,719	672	News
ABC	47,559	3.0%	yes	3,096,995	573	News
TweetLikeAGirl	45,621	2.8%	no	1,161,704	6	Humour

Table 6. Categorisation of the top 20 retweeted accounts in the 5th sub-period.

Twitter user screen name	Number of retweets	Prevalence of retweets number among the top 20 retweeted accounts	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
LeadershipNGA	77,204	11.0%	no	79,945	21,782	Automated tweet aggregator
UNICEF	64,033	9.1%	yes	2,595,866	780	(Inter)governmental organisation/NGO
onedirection	57,447	8.2%	yes	21,169,391	2	Celebrity
nytimes	39,727	5.7%	yes	11,369,883	1,338	News
WHO	38,455	5.5%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
NiallOfficial	37,551	5.4%	yes	20,389,619	1	Celebrity
nigeriantribune	36,372	5.2%	yes	17,076	8,795	News
UN	35,403	5.1%	yes	2,475,978	571	(Inter)governmental organisation/NGO
BillGates	30,325	4.3%	yes	17,357,705	37	Celebrity
FunnyVines	29,612	4.2%	no	933,952	8	Humour
TIME	29,065	4.2%	yes	5,572,684	743	News
CapcyCom	28,007	4.0%	no	4,548	26	Twitter celebrity
coldplay	26,626	3.8%	yes	13,395,712	3	Celebrity
conspiracyimage	25,918	3.7%	no	724,491	3	Humour
MSF_espana	25,786	3.7%	yes	287,304	601	(Inter)governmental organisation/NGO
UNMEER	24,876	3.6%	yes	110	1,544	(Inter)governmental organisation/NGO
RelatableQuote	24,425	3.5%	no	2,779,084	7	Humour
BBCBreaking	23,331	3.3%	yes	9,222,667	165	News
BBCAfrica	23,240	3.3%	yes	369,999	1,219	News
Independent	23,182	3.3%	yes	569,962	566	News

Appendix 7. Most retweeted tweets and the types of accounts that posted them in each of the study sub-periods.

Table 1. Top 10 most retweeted tweets and the types of accounts that posted them in the 1st sub-period.

Number	Account screen name	Tweet text	Number of retweets	Lowest number of the account followers in the dataset	Account type
1	rickygervais	****	319	7,019,229	Celebrity
2	gengelis	Madurismo? Diosdadismo? Amelichismo? Estos escuacas están es locos'ebola! Aquí lo que hay es Chavismo. Cha-vis-mo. #MásNá (Translation: Madurismo? Diosdadismo? Amelichismo? These are escuacas is locos'ebola! What we have here is Chavismo. Cha-vis-mo. #Masna)	226	2,359	Twitter celebrity
3	GarethBale22	On a scale of the Ebola virus to Adolf Hitler, how bad is Ashley Young? #ManchesterDerby	167	87,229	Humour
4	MaraWritesStuff	****	111	74,107	Celebrity
5	medschooladvice	Every med student at some point thinks they have a disease they are learning about. Relax, you don't have Ebola!	106	61,212	Twitter celebrity
6	Vice_Is_Hip	****	102	60,872	Humour
7	yotobi	Credo che tutti i creatori di siti di Gossip finiranno all'inferno, dove un enorme uomo malato di ebola gli stuprerà la faccia di violenza. (Translation: I think that all creators of Gossip sites end up in hell, where a huge man sick of ebola stuprerà the face of violence.)	78	46,95	Twitter celebrity
8	HuffingtonPost	As chimpanzee numbers plunge across Africa, scientists discover massive chimp "mega-culture" in remote Congo forest http://t.co/ufB4G0VKio	76	3,842,480	News
9	****	****	71	261	Twitter celebrity
10	****	****	68	18,505	Twitter celebrity

**** Tweet or account has been deleted.

Table 2. Top 10 most retweeted tweets and the types of accounts that posted them in the 2nd sub-period.

Number	Account screen name	Tweet text	Number of retweets	Lowest number of the account followers in the dataset	Account type
1	La_Dexterite	Le gar a payé mèche, iPhone pour elle. Au moment de Mougou la go dit "J'ai EBOLA" le gar dit "même si ta DEBORAH je vais te MOUGOU audhui" (Translation: The boy paid wick iPhone for her. When the Mougou says go "I Ebola" the boy said, "even if your DEBORAH I MOUGOU you audhui")	2,812	7,043	Twitter celebrity
2	WHO	What is #Ebola? Ebola is a severe, often fatal illness, with up to 90% of people who have it dying from it http://t.co/pxSHAodnE5 #AskEbola	860	1,173,156	(Inter)governmental organisation/NGO
3	MrVop	ไวรัสอีโบลาระบาดหนักในกินี ตายแล้ว 90+ http://t.co/F7qmsqKcbX http://t.co/voFglpJ7sV (Translation: Ebola virus outbreak in Guinea dead 90+ http://t.co/F7qmsqKcbX http://t.co/voFglpJ7sV)	857	43,683	Twitter celebrity
4	cnni	Three cities. Hundreds dead. Can the deadliest ever #Ebola virus outbreak be stopped? http://t.co/0tnezAnW3q http://t.co/PP40phFwXN	835	1,895,824	News
5	buffalogags	A: มึงเคยฟังเพลง "พระ" ของ Ebola ไหม ? B: เฮ้ย! เพลงไรวะ A: สิ่งทีฉันเป็น by MrFilm (Translation:A: I've been listening to "King" of Ebola it? B: Hey! I think music A: What I by MrFilm)	752	98,683	Twitter celebrity
6	cnnbrk	142 deaths have been reported from Ebola outbreak in Guinea and Liberia, @WHO said. http://t.co/axP3KGtVpd	744	16,128,604	News
7	NewDay	The #Ebola virus outbreak in West Africa has hit "unprecedented" proportions: http://t.co/ZUkjivieUI #5Things #NewDay http://t.co/9kEgWBGw8b	736	55,603	News
8	CNN	At least 59 people are dead in Guinea from an Ebola outbreak as the deadly hemorrhagic fever quickly spreads. http://t.co/cqqzmirUsq	719	12,264,670	News
9	UN	What is #Ebola? @WHO answer frequently asked questions here: http://t.co/djHj7RQm8h http://t.co/TxoiGyWjIP	690	2,475,978	(Inter)governmental organisation/NGO

10	FumiHawk	<p>メディアはセクハラやじばかり報道していますが、アフリカではエボラ出血熱が流行しコントロール不能状態になっています。既にギニア、シエラレオネ、リベリアで 330 人以上の死者が出ている模様です。時々はこちらも思い出してあげてください。 http://t.co/4bx81IZ09y (Translation: Media has reported only sexual harassment jeers, but it has become uncontrollable state outbreak is Ebola hemorrhagic fever in Africa. It is already seemed to Guinea, Sierra Leone, we have come up with 330 or more people dead in Liberia. Please occasionally was raised Recall also here. http://t.co/4bx81IZ09y)</p>	662	6,444	Twitter celebrity
----	----------	--	-----	-------	-------------------

**** Tweet or account has been deleted.

Table 3. Top 10 most retweeted tweets and the types of accounts that posted them in the 3rd sub-period.

Number	Account screen name	Tweet text	Number of retweets	Lowest number of the account followers in the dataset	Account type
1	emmkaff	<p>Scientists: Don't freak out about Ebola. Everyone: *Panic!*</p> <p>Scientists: Freak out about climate change. Everyone: LOL! Pass me some coal.</p>	38,633	2,175	Twitter celebrity
2	BallinInHD	<p>This soap is \$195, it better wash Ebola, wash HIV, wash Malaria, shit.... It better wash all my sins away ☹ http://t.co/OXEruuAOyj</p>	34,655	45,098	Twitter celebrity
3	CNN	<p>#BREAKING: 1st diagnosed case of #Ebola in the U.S. confirmed. CDC press conference at 5:30 pm http://t.co/t8DihVMOIn http://t.co/FIEwS30UQ1</p>	25,191	12,264,670	News
4	UrbanLePharaon	<p>"un poisson rouge opéré d'une tumeur au cerveau http://t.co/vQyLlaFRmZ http://t.co/3T3faoHcWe" en Afrique y'a Ebola et eux y font sa ☹ (Translation:"goldfish surgery for a brain tumor http://t.co/vQyLlaFRmZ http://t.co/3T3faoHcWe" there's Ebola in Africa and they are doing their ☹)</p>	13,247	25,73	Twitter celebrity
5	Shiplives	<p>Ebola has killed fewer than 5,000 people...ever.</p> <p>Measles (which some folks refuse to vaccinate kids for) killed 122,000 in 2012 alone.</p>	11,705	4,006	Twitter celebrity
6	****	****	9,695	19,657	Twitter celebrity
7	DragonflyJonez	<p>How the hell we supposed to trust Atlanta to keep ebola quarantined when Gucci Mane released 8 mixtapes from jail?</p>	9,505	37,247	Twitter celebrity

8	TheTweetOfGod	"Doctor Who Survived Ebola: 'God Saved My Life'." Yep. And I killed the hundreds of Africans who died of it, because screw them.	9,425	1,480,204	Humour
9	TweetfrmtheSole	Ebola was incurable when it was killing Africans left and right but now that two white people got it we found a "secret" cure	7,243	1,331	Twitter celebrity
10	Kojoanan	EBOLA WATCH. Please share. http://t.co/WW1dLL73F2	7,007	851	Individual

**** Tweet or account has been deleted.

Table 4. Top 10 most retweeted tweets and the types of accounts that posted them in the 4th sub-period.

Number	Account screen name	Tweet text	Number of retweets	Lowest number of the account followers in the dataset	Account type
1	Kyheim	Ebola jokes are funny cus it kills africans but cancer jokes aren't cus it kills white ladies . THIS TEA BOMB AF THO	62,739	2,406	Twitter celebrity
2	chrisbrown	****	45,923	13,672,027	Celebrity
3	****	****	39,749	11,512	Twitter celebrity
4	Ashton5SOS	The Ebola virus is super scary, hopefully it can be contained!	39,117	2,927,026	Celebrity
5	therealbanksy	Hunger kills more than Ebola but it's not considered a significant problem since rich people can't die of it.	37,213	766,937	Humour
6	jacob_binke	Ebola has been in the US for 1 day and people are already wearing masks. AIDS has been here for 55 years and people still don't use a condom	35,433	537	Individual
7	malonebarry	This disturbing André Carrilho illustration on Ebola, and how the world sees it, is both powerful and accurate.	32,614	7,25	Twitter celebrity
8	ArianaGrande	ariana grande! afraid to go to NYC because of Ebola!!!" a headline I read while casually walking down the street in New York	29,084	20,051,441	Celebrity
9	druggable	Retweet this in 15 seconds or you're gonna get Ebola	28,302	14,15	Twitter celebrity
10	B_JeffTheGod	Akon crowd surfed in a plastic bubble to avoid catching Ebola. I'm sorry but this got me weak as hell	26,817	1,737	Twitter celebrity

**** Tweet or account has been deleted.

Table 5. Top 10 most retweeted tweets and the types of accounts that posted them in the 5th sub-period.

Number	Account screen name	Tweet text	Number of retweets	Lowest number of the account followers in the dataset	Account type
1	NiallOfficial	BandAid30: We're very happy to have your support. Thanks @NiallOfficial #BandAid30 http://t.co/jT7zWnsXNK "pleasure	37,567	20,389,619	Celebrity
2	onedirection	One Direction are joining the @BandAid30 single with @edsheeran, @OfficialAdele, @elliegoulding & more!!! http://t.co/8m3myuSwgW #BandAid3	29,793	21,169,391	Celebrity
3	onedirection	Let's #StopTheSpread of Ebola together - support @SaveChildrenUK & @DECappeal and text DEC to 70000 to donate £5!	28,473	21,169,391	Celebrity
4	MittRomney	SonyPictures don't cave, fight: release @TheInterview free online globally. Ask viewers for voluntary \$5 contribution to fight #Ebola.	28	1,602,249	Celebrity
5	coldplay	To donate £5 to help @BandAid30 fight ebola, text AID to 70060 (UK only) or donate at http://t.co/lplgPg2zdP (worldwide).	24,755	13,395,712	Celebrity
6	conspiracyimage	EBOLA, EBOLA, EBOLA...everyday, for weeks and then all of a sudden ...NOTHING!. http://t.co/iukM1emE07	22,459	724,491	Humour
7	AllergyKidsDoc	Actual discussion: Parent "I want #Ebola vaccine for my child" Doc "There isn't one, but we have #flushot" Parent "We don't believe in that"	16,81	1,9	Twitter celebrity
8	Vevo_UK	Buying @BandAid30 goes directly to fighting Ebola. Watch the video now which includes @onedirection + @Ollyofficial! http://t.co/IEYtmqcPmY	15,832	174,016	Other verified account
9	stinsonsays	*Ebola joke* everyone laughs *AIDS joke* everyone laughs *Cancer joke* wtf is wrong with you? people die of Cancer! -our stupid society"	15,829	724,517	Humour
10	RelatableQuote	all I hear in school. -ebola -yeet -bruhh -whats 9 + 10 -21 -ebola -tacos -my anaconda don't -ebola -bout a week ago -alexfromtarget	14,814	2,779,084	Humour

**** An account has been deleted or suspended or a tweet has been deleted

Appendix 8. Features of the top 20 mentioned accounts over the whole study period and each of the sub-periods used for their categorisation.

Table 1. Categorisation of the top 20 mentioned accounts in the full study period.

Twitter user screen name	Number of tweets mentioning the Twitter user	Prevalence of tweets mentioning the account among the top 20	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
YouTube	20,9896	22.5%	yes	44,080,431	4	Other verified account
ShareThis	67,002	7.2%	no	1,712,726	3	Twitter celebrity
233liveOnline	64,613	6.9%	no	1	64,559	Automated tweets aggregator
nytimes	63,820	6.8%	yes	11,369,883	1,338	News
WHO	60,056	6.4%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
naijaXclusive	56,433	6.1%	no	108,539	231	Twitter celebrity
CNN	36,879	4.0%	yes	12,264,670	157	News
CDCgov	36,581	3.9%	yes	292,595	624	(Inter)governmental organisation/NGO
AP	36,206	3.9%	yes	3,096,719	672	News
washingtonpost	31,294	3.4%	yes	2,482,703	609	News
BarackObama	31,116	3.3%	yes	47,042,588	5	Celebrity
FoxNews	29,723	3.2%	yes	3,770,972	391	News
YahooNews	29,396	3.2%	yes	762,887	998	News
Infotainment9ja	28,715	3.1%	no	1,965	576	Twitter celebrity
cnnbrk	28,558	3.1%	yes	16,128,604	260	News
Reuters	27,684	3.0%	yes	4,097,821	314	News
guardian	25,666	2.8%	yes	2,081,092	773	News
MailOnline	24,511	2.6%	yes	547,834	209	News
NBCNews	22,207	2.4%	yes	1,397,688	487	News
BreitbartNews	22,104	2.4%	no	157,808	558	Twitter celebrity

Table 2. Categorisation of the top 20 mentioned accounts in the 1st sub-period.

Twitter user screen name	Number of tweets mentioning the Twitter user	Prevalence of tweets mentioning the account among the top 20	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
FluTrackers	955	36.1%	no	3,249	6,48	Infectious disease tracking
YouTube	676	25.6%	yes	44,080,431	4	Other verified account
ironorehopper	166	6.3%	no	333	6,638	Automated tweet aggregator
Diane_Morin	109	4.1%	n/a	n/a	0	n/a
lilithlela	107	4.0%	no	4,435	412	Twitter celebrity
GertvanderHoek	88	3.3%	no	659	1,332	Infectious disease tracker
rickygervais	63	2.4%	yes	7,019,229	1	Celebrity
mixin1771	62	2.3%	no	26	1	Other non-verified
washingtonpost	51	1.9%	yes	2,482,703	609	News
HuffPostGreen	50	1.9%	yes	221,153	124	News
WHO	49	1.9%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
midnight	45	1.7%	yes	290.390	6	Other verified
****	41	1.6%	no	198	229	Individual
RonanKelly13	37	1.4%	no	374	2,527	Infectious disease tracking
DeezerThailand	27	1.0%	no	39	2	Other non-verified
Fla_Medic	27	1.0%	no	1,675	317	Twitter celebrity
OUPAcademic	25	1.0%	no	35,328	73	Twitter celebrity
guardian	23	0.9%	yes	2,081,092	773	News
****	23	0.9%	no	652	71	Individual
rthklocal_e	22	0.8%	no	49	33	Other non-verified

**** Account has been deleted.

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).

Table 3. Categorisation of the top 20 mentioned accounts in the 2nd sub-period.

Twitter user screen name	Number of tweets mentioning the Twitter user	Prevalence of tweets mentioning the account among the top 20	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
ShareThis	5,790	22.8%	no	1,712,726	3	Twitter celebrity
WHO	4,187	16.5%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
YouTube	3,643	14.4%	yes	44,080,431	4	Other verified account
YahooNews	1,259	5.0%	yes	762,887	998	News
Reuters	1,201	4.7%	yes	4,097,821	314	News
lemondefr	1,027	4.1%	yes	2,141,488	246	News
AJEnglish	942	3.7%	yes	1,904,039	232	News
nprnews	745	2.9%	yes	2,301,095	135	News
FluTrackers	709	2.8%	no	3,249	6,48	Infectious disease tracking
guardian	666	2.6%	yes	2,081,092	773	News
AP	600	2.4%	yes	3,096,719	672	News
CNN	595	2.4%	yes	12,264,670	157	News
scoopit	547	2.2%	n/a	n/a	0	n/a
BloombergNews	545	2.2%	yes	1,393,931	318	News
BBCWorld	538	2.1%	yes	5,901,375	396	News
HuffPostUKPol	520	2.1%	yes	17,976	25	News
NBCNews	496	2.0%	yes	1,397,688	487	News
mashable	484	1.9%	yes	3,903,494	362	News
washingtonpost	454	1.8%	yes	2,482,703	609	News
411Vibes	429	1.7%	no	33,468	113	Twitter celebrity

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).

Table 4. Categorisation of the top 20 mentioned accounts in the 3rd sub-period.

Twitter user screen name	Number of tweets mentioning the Twitter user	Prevalence of tweets mentioning the account among the top 20	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
YouTube	43,585	14.9%	yes	44,080,431	4	Other verified account
naijaXclusive	32,461	11.1%	no	108,539	231	Twitter celebrity
WHO	25,334	8.6%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
Infotainment9ja	23,204	7.9%	no	1,965	576	Twitter celebrity
nytimes	21,493	7.3%	yes	11,369,883	1,338	News
ShareThis	19,366	6.6%	no	1,712,726	3	Twitter celebrity
TheTrentOnline	15,133	5.2%	no	35,258	2,27	Twitter celebrity
411Vibes	13,455	4.6%	no	33,468	113	Twitter celebrity
Reuters	10,393	3.5%	yes	4,097,821	314	News
NigeriaNewsdesk	10,151	3.5%	yes	520,795	6,719	News
AP	9,812	3.4%	yes	3,096,719	672	News
washingtonpost	9,788	3.3%	yes	2,482,703	609	News
YahooNews	8,978	3.1%	yes	762,887	998	News
GoldRushNigeria	7,707	2.6%	no	1,777	528	Twitter celebrity
Bayyou11DOTcoM	7,607	2.6%	no	2,265	291	Twitter celebrity
cnnbrk	7,289	2.5%	yes	16,128,604	260	News
Yeremiix3	7,223	2.5%	no	59	98	Other non-verified
CDCgov	6,906	2.4%	yes	292,595	624	(Inter)governmental organisation/NGO
CNN	6,876	2.3%	yes	12,264,670	157	News
AKPRAISE	6,607	2.3%	no	42,116	654	Twitter celebrity

Table 5. Categorisation of the top 20 mentioned accounts in the 4th sub-period.

Twitter user screen name	Number of tweets mentioning the Twitter user	Prevalence of tweets mentioning the account among the top 20	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
YouTube	90,607	90.6%	yes	44,080,431	4	Other verified account
nytimes	31,898	7.4%	yes	11,369,883	1,338	News
CNN	25,348	5.9%	yes	12,264,670	157	News
CDCgov	24,665	5.8%	yes	292,595	624	(Inter)governmental organisation/NGO
BarackObama	22,380	5.2%	yes	47,042,588	5	Celebrity
FoxNews	20,972	4.9%	yes	3,770,972	391	News
naijaXclusive	20,629	4.8%	no	108,539	231	Twitter celebrity
ShareThis	19,739	4.6%	no	1,712,726	3	Twitter celebrity
AP	18,987	4.4%	yes	3,096,719	672	News
cnnbrk	18,335	4.3%	yes	16,128,604	260	News
washingtonpost	16,701	3.9%	yes	2,482,703	609	News
BreitbartNews	15,064	3.5%	no	157,808	558	Twitter celebrity
MailOnline	14,566	3.4%	yes	547,834	209	News
WHO	14,176	3.3%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
YahooNews	14,067	3.3%	yes	762,887	998	News
NBCNews	13,617	3.2%	yes	1,397,688	487	News
ABC	12,776	3.0%	yes	3,096,995	573	News
WorldStarFunny	12,044	2.8%	no	865,369	29	Humour
guardian	11,275	2.6%	yes	2,081,092	773	News
USATODAY	11,098	2.6%	yes	1,015,727	352	News

Table 6. Categorisation of the top 20 mentioned accounts in the 5th sub-period.

Twitter user screen name	Number of tweets mentioning the Twitter user	Prevalence of tweets mentioning the account among the top 20	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
YouTube	71,385	71.4%	yes	44,080,431	4	Other verified account
233liveOnline	52,181	52.2%	no	1	64,559	Automated tweet aggregator
ShareThis	22,095	8.0%	no	1,712,726	3	Twitter celebrity
LeadershipNGA	17,358	6.3%	no	79,945	21,782	Automated tweet aggregator
WHO	16,310	5.9%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
nytimes	10,180	3.7%	yes	11,369,883	1,338	News
TIME	9,697	3.5%	yes	5,572,684	743	News
guardian	7,405	2.7%	yes	2,081,092	773	News
CanaisTvOnline	7,213	2.6%	no	22,437	385	Twitter celebrity
iam_Davido	7,108	2.6%	yes	856,359	30	Celebrity
kamalajiboye	6,983	2.5%	no	6,229	11	Twitter celebrity
AP	6,807	2.5%	yes	3,096,719	672	News
Reuters	6,148	2.2%	yes	4,097,821	314	News
R2Pindex	5,895	2.1%	no	0	10,917	Automated tweet aggregator
UN	5,496	2.0%	yes	2,457,978	571	(Inter)governmental organisation/NGO
nigeriantribune	5,176	1.9%	yes	17,076	8,795	News
YahooNews	5,092	1.8%	yes	762,887	998	News
nprnews	5,087	1.8%	yes	2,301,095	135	News
UNMEER	5,031	1.8%	yes	110	1,544	(Inter)governmental organisation/NGO
UNICEF	4,867	1.8%	yes	2,595,866	780	(Inter)governmental organisation/NGO

Appendix 9. Features of the top 20 accounts that received the highest number of replies over the whole study period and each of the sub-periods used for their categorisation.

Table 1. Categorisation of the top 20 accounts that received the highest number of replies in the full study period.

Twitter user screen name	Number of tweets containing replies	Prevalence of tweets containing replies among the top 20 with the highest number of replies	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
cnnbrk	12469	8.1%	yes	16,128,604	260	News
WHO	12160	7.9%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
nytimes	11138	7.2%	yes	11,369,883	1,338	News
R2Pindex	10909	7.1%	no	0	10,917	Automated tweet aggregator
WorldStarFunny	10367	6.7%	no	865,369	29	Humour
CNN	9754	6.3%	yes	12,264,670	157	News
FoxNews	8445	5.5%	yes	3,770,972	391	News
BarackObama	8418	5.4%	yes	47,042,588	5	Celebrity
AP	7423	4.8%	yes	3,096,719	672	News
CDCgov	7199	4.7%	yes	292,595	624	(Inter)governmental organisation/NGO
FunnyPicsDepot	6876	4.4%	no	1,178,130	27	Humour
Dory	6348	4.1%	no	669,753	39	Humour
FunnyVines	5778	3.7%	no	933,952	8	Humour
ItsYaVirusEBOLA	5749	3.7%	no	357	77	Humour
TIME	5529	3.6%	yes	5,572,684	743	News
****	5413	3.5%	no	n/a	0	n/a
WhiteHouse	5306	3.4%	yes	5,264,927	116	(Inter)governmental organisation/NGO
WORIDSTARHIPH OP	5246	3.4%	no	242,816	38	Humour
realDonaldTrump	5224	3.4%	yes	2,645,768	93	Celebrity
NBCNews	5076	3.3%	yes	1,397,688	487	News

**** Account has been deleted.

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).

Table 2. Categorisation of the top 20 accounts that received the highest number of replies in the 1st sub-period.

Twitter user screen name	Number of tweets containing replies	Prevalence of tweets containing replies among the top 20 with the highest number of replies	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
rickygervais	45	17.1%	yes	7,019,229	1	Celebrity
WHO	26	9.9%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
FluTrackers	24	9.1%	no	3,249	648	Infectious disease tracking
GertvanderHoek	16	6.1%	no	659	1,332	Infectious disease tracker
MaraWritesStuff	15	5.7%	yes	74,107	18	Celebrity
AristeguiOnline	14	5.3%	yes	2,11,385	226	News
****	13	4.9%	no	652	71	Individual
AverageRejcect	12	4.6%	no	52	70	Other non-verified
MackayIM	12	4.6%	no	346	3,164	Automated tweet aggregator
****	12	4.6%	no	50	0	Individual
Atari2600hero	8	3.0%	no	72	63	Other non-verified
****	8	3.0%	n/a	n/a	0	n/a
****	8	3.0%	n/a	n/a	0	n/a
ebolasound	8	3.0%	no	2,835	753	Twitter celebrity
****	8	3.0%	n/a	n/a	0	n/a
midnight	8	3.0%	yes	290,390	6	Other verified
Ebolajong	7	2.7%	no	501	5	Other non-verified
RonanKelly13	7	2.7%	no	374	2,527	Infectious disease tracking
er_noir_sfina	7	2.7%	no	3,139	152	Twitter celebrity
****	6	2.3%	no	788	11	Individual

**** Account has been deleted.

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).

Table 3. Categorisation of the top 20 accounts that received the highest number of replies in the 2nd sub-period.

Twitter user screen name	Number of tweets containing replies	Prevalence of tweets containing replies among the top 20 with the highest number of replies	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WHO	1,293	36.7%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
HaertIG	183	5.2%	no	1,286	867	Twitter celebrity
BBCWorld	177	5.0%	yes	5,901,375	396	News
AirFranceFR	142	4.0%	yes	356,501	4	Other verified
matteosalvinimi	139	3.9%	yes	35,444	17	Celebrity
cnnbrk	137	3.9%	yes	16,128,604	260	News
AP	130	3.7%	yes	3,096,719	672	News
BBCAfrica	129	3.7%	yes	369,999	1,219	News
cnni	123	3.5%	yes	1,895,824	186	News
DRUDGE_REPORT	116	3.3%	no	647,372	926	Automated tweet aggregator
airfrance	112	3.2%	yes	104,022	2	Other verified
CNN	104	3.0%	yes	12,264,670	157	News
MackayIM	103	2.9%	no	346	3,164	Automated tweet aggregator
SakhalinTribune	102	2.9%	no	0	8	Other non-verified
****	97	2.8%	n/a	n/a	0	n/a
Reuters	92	2.6%	yes	4,097,821	314	News
AJEnglish	90	2.6%	yes	1,904,039	232	News
****	90	2.6%	n/a	n/a	0	n/a
lemondefr	86	2.4%	yes	2,141,488	246	News
drsanjaygupta	81	2.3%	yes	1,803,753	59	Celebrity

**** Account has been deleted.

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).

Table 4. Categorisation of the top 20 accounts that received the highest number of replies in the 3rd sub-period.

Twitter user screen name	Number of tweets containing replies	Prevalence of tweets containing replies among the top 20 with the highest number of replies	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WHO	5,420	13.6%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
cnnbrk	3,071	7.7%	yes	16,128,604	260	News
R2Pindex	2,947	7.4%	no	0	10,917	Automated tweet aggregator
nytimes	2,945	7.4%	yes	11,369,883	1,338	News
****	2,417	6.1%	n/a	n/a	0	n/a
NigeriaNewsdesk	2,339	5.9%	yes	520,795	6,719	News
****	1,949	4.9%	n/a	n/a	0	n/a
AP	1,915	4.8%	yes	3,096,719	672	News
olofofounded	1,847	4.6%	no	153	332	Other non-verified
realDonaldTrump	1,637	4.1%	yes	2,645,768	93	Celebrity
EbolaAlert	1,551	3.9%	no	4	5,592	Automated tweet aggregator
BarackObama	1,494	3.8%	yes	47,042,588	5	Celebrity
CNN	1,481	3.7%	yes	12,264,670	157	News
CDCgov	1,464	3.7%	yes	292,595	624	(Inter)governmental organisation/NGO
Gidi_Traffic	1,341	3.4%	no	112,768	937	Twitter celebrity
BBCBreaking	1,332	3.3%	yes	9,222,667	165	News
washingtonpost	1,244	3.1%	yes	2,482,703	609	News
BBCAfrica	1,195	3.0%	yes	369,999	1,219	News
****	1,118	2.8%	n/a	n/a	0	n/a
ABC	1,090	2.7%	yes	3,096,995	573	News

**** Account has been deleted.

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).

Table 5. Categorisation of the top 20 accounts that received the highest number of replies in the 4th sub-period.

Twitter user screen name	Number of tweets containing replies	Prevalence of tweets containing replies among the top 20 with the highest number of replies	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
WorldStarFunny	8,779	8.5%	no	865,369	29	Humour
cnnbrk	8,237	8.0%	yes	16,128,604	260	News
CNN	7,357	7.1%	yes	12,264,670	157	News
FoxNews	6,571	6.4%	yes	3,770,972	391	News
nytimes	6,479	6.3%	yes	11,369,883	1,338	News
Dory	6,204	6.0%	no	669,753	39	Humour
BarackObama	6,058	5.9%	yes	47,042,588	5	Celebrity
FunnyPicsDepot	6,024	5.8%	no	1,178,130	27	Humour
ItsYaVirusEBOLA	5,749	5.6%	no	357	77	Humour
CDCgov	5,169	5.0%	yes	292,595	624	(Inter)governmental organisation/NGO
AP	4,430	4.3%	yes	3,096,719	672	News
WhiteHouse	4,272	4.1%	yes	5,264,927	116	(Inter)governmental organisation/NGO
sorryimalex	3,899	3.8%	no	258,248	38	Twitter celebrity
NBCNews	3,602	3.5%	yes	1,397,688	487	News
TweetLikeAGirl	3,583	3.5%	no	1,161,704	6	Humour
WORIDSTARHIPHOP	3,561	3.4%	no	242,816	38	Humour
YaBoyEbola	3,554	3.4%	no	0	44	Humour
chrisbrown	3,446	3.3%	yes	13,672,027	1	Celebrity
realDonaldTrump	3,299	3.3%	yes	2,645,768	93	Celebrity
ABC	3,160	3.1%	yes	3,096,995	573	News

Table 6. Categorisation of the top 20 accounts that received the highest number of replies in the 5th sub-period.

Twitter user screen name	Number of tweets containing replies	Prevalence of tweets containing replies among the top 20 with the highest number of replies	Verification by Twitter	The lowest number of account followers in the dataset	Number of tweets in the dataset	Account type
R2Pindex	5,894	14.7%	no	0	10,917	Automated tweet Aggregator
****	3,884	9.7%	n/a	n/a	0	n/a
WHO	2,714	6.8%	yes	1,173,156	1,781	(Inter)governmental organisation/NGO
FunnyVines	2,663	6.7%	no	933,952	8	Humour
KingRyin	2,608	6.5%	no	36,034	6	Twitter celebrity
TIME	1,928	4.8%	yes	5,572,684	743	News
****	1,826	4.6%	n/a	n/a	0	n/a
KTHopkins	1,798	4.5%	yes	144,353	41	Celebrity
nytimes	1,661	4.2%	yes	11,369,883	1,338	News
twaimz	1,648	4.1%	yes	74,926	13	Celebrity
UN	1,613	4.0%	yes	2,457,978	571	(Inter)governmental organisation/NGO
WorldStarFunny	1,577	3.9%	no	865,369	29	Humour
WORIDSTARHIPH OP	1,567	3.9%	no	242,816	38	Humour
LeadershipNGA	1,335	3.3%	no	79,945	21782	Automated tweet aggregator
conspiracyimage	1,274	3.2%	no	724,491	3	Humour
UNICEF	1,262	3.2%	yes	2,595,866	780	(Inter)governmental organisation/NGO
WeLoveRobDyrdek	1,244	3.1%	no	463,861	34	Humour
pewdiepie	1,214	3.0%	yes	4,083,426	2	Celebrity
UNMEER	1,194	3.0%	yes	110	1,544	(Inter)governmental organisation/NGO
UberFacts	1,170	2.9%	yes	6,786,919	36	Other verified

**** Account has been deleted.

n/a means that it was impossible to categorise the account due to the absence of needed information (they did not post any tweets included in the dataset).