Purdue University
Purdue e-Pubs

Open Access Theses
Theses and Dissertations

12-2016

An event detection approach based on Twitter hashtags

Shih-Feng Yang
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_theses

Recommended Citation

Yang, Shih-Feng, "An event detection approach based on Twitter hashtags" (2016). *Open Access Theses*. 909.
https://docs.lib.purdue.edu/open_access_theses/909

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

AN EVENT DETECTION APPROACH
BASED ON TWITTER HASHTAGS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Shih-Feng Yang

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 2016

Purdue University

West Lafayette, Indiana

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Shih-Feng Yang

Entitled

AN EVENT DETECTION APPROACH BASED ON TWITTER HASHTAGS

For the degree of Master of Science

Is approved by the final examining committee:

Julia Taylor
Chair

John Springer

Brandeis Marshall

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Julia Taylor

Approved by: Jeffrey Whitten 5/19/2016

Head of the Departmental Graduate Program Date

TABLE OF CONTENTS

Page

LIST OF FIGURES iv
LIST OF TABLES v
LIST OF ABBREVIATIONS vi
GLOSSARY vii
ABSTRACT viii
CHAPTER 1. INTRODUCTION 1
  1.1 Background 1
  1.2 Significance 2
  1.3 Statement of purpose 3
  1.4 Research question 3
  1.5 Assumptions 3
  1.6 Limitations 4
CHAPTER 2. LITERATURE REVIEW 5
  2.1 Introduction 5
  2.2 Event detection 5
  2.3 Hashtag analysis 12
  2.4 Summary 18
CHAPTER 3. METHODOLOGY 19
  3.1 Introduction 19
  3.2 Data collection and preprocessing 21
  3.3 Hashtag clustering 24
    3.3.1 Hashtag and event representations 24
    3.3.2 K-means clustering 26
    3.3.3 STREAMCUBE clustering 28

Page

3.4 Data analysis 29
CHAPTER 4. EXPERIMENTS 33
4.1. Data Collection 33
4.2. Compare K-means to STREAMCUBE (STREAMCUBE as the ground truth)...34
4.3. Compare K-means to STREAMCUBE (human serving as the ground truth)....38
4.4. Find better K values for K-means (human serving as the ground truth) 42
CHAPTER 5. CONCLUSION 45
5.1. Conclusion 45
5.2. Recommendations for future studies 46
LIST OF REFERENCES 48
APPENDICES
APPENDIX A: NUMBER OF ELEMENTS IN THE CLUSTERS 52
APPENDIX B: TWEETS MANUALLY LABELED BY A HUMAN EXPERT....56

LIST OF FIGURES

Page

Figure 3.1. Workflow overview 20
Figure 3.2. System architecture 21
Figure 3.3. Tweet preprocessing 22
Figure 3.4. HASHTAG-CLUSTER-STATIC algorithm (Feng et al., 2015)....................29
Figure 3.5. NEAREST-NEIGHBOR algorithm (Feng et al., 2015) 29
Figure 3.6. The equation of the Purity (1) 30
Figure 3.7. The equation of the Purity (2) 31
Figure 3.8. The equation of NMI (1) 31
Figure 3.9. The equation of NMI (2) 32
Figure 4.1. The distribution of the human labeled tweets....................39
Figure 4.2. The line graph of Purity scores for different number of clusters....................44
Figure 4.3. The line graph of NMI scores for different number of clusters....................44

LIST OF TABLES

Page

Table 4.1. The NMI and Purity scores of every 6 hours 36
Table 4.2. The NMI and Purity scores of every 12 hours 37
Table 4.3. The NMI and Purity scores of every 24 hours 37
Table 4.4. The summary of table 4.1, 4.2 and 4.3 38
Table 4.5. Hashtags of the top 15 large clusters generated by the K-means approach 40
Table 4.6. Hashtags of the top 15 large clusters generated by STREAMCUBE 41
Table 4.7. The performance comparison based on human-labeled categories...................42
Table 4.8. The performance of the K-means approach with different Ks..........................43

LIST OF ABBREVIATIONS

API - Application Program Interface

NMI - Normalized Mutual Information

GLOSSARY

Event detection - The methodology of utilizing social networks to determine whether special events, such as holidays, sport games, earthquakes, crimes, are happening.

Hashtag - A word or phrase proceeded by “#”.

Tweet - An 140-character message posted by Twitter users.

Twitter - An online social networking service.

ABSTRACT

Yang, Shih-Feng. M.S., Purdue University, December 2016. An Event Detection Approach Based On Twitter Hashtags. Major Professor: Julia Taylor Rayz.

Twitter is one of the most popular microblogging services in the world. The great amount of information made Twitter an important information channel for people to know and share news. Hashtag is a popular feature when people use Twitter. It can be taken as human labeled information and is useful for people to identify the topic of a tweet. Many researchers have proposed event-detection approaches that can monitor Twitter data and determine whether special events, such as accidents, extreme weather, earthquakes, or crimes, are happening. Although many approaches considered hashtag as one of their features, few of them explicitly focused on the effectiveness of using hashtag on event detection. In this study, we proposed an event detection approach that utilizes hashtags in tweets. We adopted the feature extraction used in STREAMCUBE (Feng et al., 2015) and applied a clustering K-means approach (Lloyd, 1982) to it. The experiments were conducted on 20,514 tweets with 8,616 hashtags collected between November 13, 2015 and November 17, 2015 with general topic of the Paris Attacks. A randomly sampled subset of 200 tweets was also manually labeled by a human subject to verify the approach. Based on the collected tweets, we demonstrated that the K-means

approach could perform better than STREAMCUBE in the clustering results. Also, we discussed how to set the K values for the K-means approach to lead to a better clustering performance.

CHAPTER 1. INTRODUCTION

1.1 Background

Twitter is one of the most popular microblogging services in the world. There are more than 500 million Twitter posts (i.e., tweets) generated per day and around 200 billion per year. The great amount of information made Twitter an important information channel for people to know and share news. Twitter has several characteristics that distinguish it from news web sites and other information channels (Li et al., 2012). First, tweets are created in real-time. For example, a tweet related to a tornado might be written one minute after a user witnessed a tornado was formed. The information could be spread even faster than TV broadcasts. Second, tweets contain any kinds of information shared by people. When people see gunfire, earthquake or other events, every witness can share his observations and pictures immediately. The information could be helpful to evaluate the actual situation of the events. Third, tweets contain geolocation information. By monitoring tweets about crime events in a specific location, some crimes could be detected immediately.

Hashtag is a popular feature when people use Twitter. A hashtag is a word or phrase proceeded by "#", and is used to identify messages on a specific topic (Feng et al., 2015). For example, "#ParisAttacks" can be used to indicate the terrorist attacks happened in evening of November 13, 2015. It is an important feature for researchers to identify the topic of a tweet.

Recently, many organizations have utilized crowdsourced information from Twitter to detect natural disaster events (Middleton et al., 2014). Utilizing tweets in event detection helps researchers understand further about an event. The goal of this research was to design a clustering approach that utilized hashtag and to analyze the similarity and difference between the clusters generated by other existing approaches and one proposed. We also surveyed the state-of-the-art event detection approaches based on Twitter data, especially the applications that utilized hashtag.

1.2 Significance

Many researchers proposed event-detection approaches that monitored Twitter data and determined whether special events, such as accidents, extreme weather, earthquakes, or crimes, were happening by analyzing the data on social networks. Data mining techniques related to clustering, classification, and text mining techniques were wildly used in this topic. Although many of them considered hashtag as one of their features, few of them explicitly focused on the effectiveness of hashtag on event detection. In order to filled this gap and make the best use of hashtag on Twitter clustering, we discussed the effectiveness of hashtag and proposed an event detection approach in this study. Moreover, we describe complete experiments to evaluate the the clustering quality of our approach as well as the comparisons with other clustering approaches.

1.3 Statement of purpose

The purpose of this research was to propose an event detection approach that utilized hashtags in tweets. We adopted the feature extraction used in the approach of Feng et al. (2015) and the K-means (Lloyd, 1982) clustering method. We surveyed the background of current related works and explain the performance differences between those works and the proposed approach. According to the experimental results, we discussed the possible improvements for the current research of event detection using Twitter.

1.4 Research question

The question central to this study was:

1. Can the state-of-the-art approaches of event detection using Twitter hashtag be improved by introducing K-means clustering?

1.5 Assumptions

The assumptions of this research included:

1. Authors of the collected tweets were assumed to write their true observations and did not intentionally fabricate false content.

2. The Twitter APIs were functional and could return correct information of public tweets during November 13th to November 17th, 2015.
3. The third-party K-means libraries used in our implementation had no significant defects that may lead to incorrect experiment results.

1.6 Limitations

The limitations of this research included:

1. In order to compare the performance with related works, we contacted the authors of Feng et al. (2015) for accessing their source codes and datasets. However, we were not able to receive the feedbacks from the authors. We implemented this approach according to the steps provided in these articles.
2. Because Twitter APIs only allowed developers to access public tweets, our data crawler only collected public tweets for experiments.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

The literature review focused on reviewing past research of the following fields. Many articles of event detection utilized Twitter data to determine whether special events, such as holidays, sport games, earthquakes, crimes, are happening. The research of hashtag analysis made use of hashtags in tweets to determine the sentiments, preferences and topics of tweets.

2.2 Event detection

Middleton et al. (2014) presented a real-time crisis-mapping platform. The researchers implemented an offline service including data-extraction tools for extracting geospatial data. The proposed system performed data preprocessing such as geocoding for every location address. Their real-time service had a Twitter crawler and a location-extraction module for efficient real-time matching. Their system contained a parallel geospatial clustering service to continuously cluster spatial areas of high activity. They performed two case studies by comparing their crisis map with the damaged area in Oklahoma's 2013 tornado and in Hurricane Sandy (October 2012). They compared their maps to the storm-surge map from the official post-event impact assessment produced by the US National Geospatial Agency (NGA). The researchers segmented the maps into 8 x 8 grid and compared each grid cells to the growth truth map. True positives were reported for any cell that has both a tweeted location and some storm-surge activity on the US NGA impact assessment map. The maps were computed using a high/medium/low threshold setting for the allowed deviation of simple moving average from baseline (dev_sma). When increasing dev_sma, the number of places and streets on the maps will be fewer, so basically a higher precision and lower recall will be obtained. The researchers calculated precision, recall, and F1 measures as their performance metrics. When dev_sma was greater than 0.1, both of their case studies obtained higher than 90% precision. However, the recall were both lower than 30%.

Walther and Kaisser (2013) proposed an algorithm for geo-spatial event detection on social media streams. The researchers monitored all Twitter posts that were in a given geographic region and identified the locations that showed a high amount of activities as event candidates. The researchers extracted textual features and other attributes from the event candidates, and used a classification component to make a binary decision of whether the candidate was an event. In the experiments, the researchers collected 1000 candidate events identified by their system. The candidate events were then manually labeled as real world events or not. The researchers than respectively adopted Naive Bayes, Multilayer Perceptron and pruned C4.5 decision tree as their classifier. In the experiments, the pruned C4.5 decision tree performed best and achieved a precision of 85.8% and a recall of 85.6%. The second experiment is to compare the difference between employing only textual features, employing only other features and employing all features. The results showed employing all features is the best, employing only textual features is the second, and employing other features is the last.

Hua et al. (2013) presented a semi-supervised system, STED, that can detect target types of events for users in Twitter. The researchers extracted action words and named entities from news articles as candidate query words and labeled tweets that contain these words. They further used social-ties terms (i.e., Mentions(@), Retweets(RT), and Hashtag(#)) and popular terms to model the label propagating to extend their label dataset. Finally, they performed text classification and location estimation based on the automatically labeled tweets. In their experiment, STED reacted to real events much faster than traditional news sources and achieved 72% precision and 74% recall.

Ritter et al. (2012) proposed an open-domain event-extraction and categorization system, which was a scalable and open-domain approach to extracting and categorizing events from status messages. Their approach aimed to discover important event categories and to classify extracted events based on latent variable models. In this research, annotated examples were not needed when classifying aggregate events. The researchers discovered event types that matched the aggregate events in an unsupervised manner. In the experiments, their work outperformed the supervised baseline by 14% because its ability to leverage large quantities of unlabeled data.

Li et al. (2012) proposed a system, TEDAS, to detect new events, to analyze the spatial and temporal patterns of an event, and to identify the importance of events. The research focused on crime and disaster related events (CDE). The researchers developed a set of efficient CDE-based crawler, classifiers, rankers and a prediction module to predict the event locations from the Twitter data. In their classifiers, they defined several Twitter-based features and CDE-specific features, and then trained a classification model to identify CDE-related tweets. By combing both Twitter-based and CDE-specific features, the researchers achieved accuracy of 80%. In their rankers, they aimed to identify important CDEs in a learning-to-rank approach that considered content features, user features and usage features. Finally, they used a linear regression model to estimate the importance of each CDE, but the performance results weren't described in this study.

Mathioudakis and Koudas (2010) presented TwitterMonitor that performs trend detection over the Twitter stream. The researchers detected the burst keywords by processing tweets with their one-pass real-time algorithm based on queuing theory. They also filtered the spurious burst and spam words. They then grouped the burst keywords to generate possible trends. The researchers implemented a front-end interface that enabled users to rank the generated trends along with a short description. However, a performance analysis was not provided in this study.

Sakaki et al. (2010) investigated the real-time interaction of events, such as earthquakes, on Twitter, proposed an algorithm to monitor tweets, and detected a target event. Each post on Twitter was classified as positive or negative by a semantic analysis in order to determine if it was truly referring to an actual earthquake occurrence. Support Vector Machine method was applied for the classification method in the semantic analysis. The main idea of this approach was taking every Twitter user as a sensor. The researchers then transformed the problem into an event-based problem on sensory observations. In their experiments, they detected 96% of earthquakes that were stronger than Japan Meteorological Agency (JMA) seismic intensity scale 3. Moreover, their average response time was within 20 seconds after an earthquake occurred, which was faster than the average response time (6 minutes) of JMA announcement.

Lee and Sumiya (2010) proposed a Twitter-based geo-social event detection system by measuring geographical regularities of crowd behaviors for Twitter. In this study, the researchers built geographical regularities deduced from the usual behavior patterns of crowds with geo-tagged microblogs. First, they proposed the Region-of-Interests (RoIs) factor based on three indicators #Tweets, #Crowd, and #MovCrowd. Second, in order to configure RoIs, they used K-means clustering method to partition tweets based on their geographical occurrences. Third, they estimated geographical regularity of local crowd behaviors. Finally, they detected “usual” geographical areas based on the three indicators and detected unusual areas as the target events. In their experiment, the researchers tried to detect the geographical areas of 15 town festivals held in Japan during July 17th to 19th, 2010. The recall achieved 87% but the precision was 1.8%, which meant their approach detected the geographical areas of the town festivals, however, it might detected many other unknown events irrelevant to the town festivals as well.

Cataldi et al. (2010) proposed a topic detection technique that retrieved the most recent topics expressed by the community in a real-time manner. The researchers extracted a set of terms from tweets and modeled each term's life cycle according to an aging theory used to score the terms. They intended to discover the emerging terms that frequently occurred in the specified time interval but relatively rare in the past. The researchers considered social relationships in the user network to quantify the importance of each analyzed content, then formalized a keyword-based topic graph that connected the emerging terms with their co-occurrent terms. After highlighting the recent hot terms, these terms were used to select related topics. These topics were taken as the most recent ones expressed by the community. The researchers provided case studies to demonstrate the effectiveness of their approach. They retrieved topics based on the five most emerging terms detected on April 15th, 2010 and showed those terms were relevant to specific news events reported in professional news articles. However, there is no comprehensive analysis about the accuracy or precision provided in this study.

Techniques of data visualization enable human to browse a large collection of tweets using a timeline-based display that highlights peaks of high tweet activity. Marcus et al. (2011) proposed TwitInfo to aggregate and visualize tweets for event exploration. The researchers designed a streaming algorithm that automatically discovers peaks of high tweet activities and labels them using text from the tweets. They also adopted a

Naïve Bayes classifier to detect the sentiments in tweets. Finally, they developed timelines, maps and pie charts for users to monitor events and explored further via geolocation, sentiment, and popular URLs. The researchers evaluated their performance by detecting different types of event. The precision was 14% to 95% and the recall was 77% to 100% according to different types of event.

Ko et al. (2014) proposed to establish an event detection approach for a large amount data with high complexity. The research was essential to integrate techniques for data analysis, visualization, and exploration. The researchers analyzed high-dimensional, multivariate network data by visualization technology. In this research, they introduced two visualization components, Petal and Thread, to effectively present a large amount of network data including multi-attribute vectors. The researchers performed case studies to evaluate the petal and thread designs by using flight delay network data from the top 50 airports. The data included the delay information of airport with five different delay types. The researchers used their components to display the five delay types as well as the delay time happened in every airport. They recruited 30 participants from various majors at the author's university and asked them to recognize the longest delay type reported in the 50 airports. The researchers first set up five hypotheses for the petal design. They evaluated the accuracy and the time that a participant spent to complete specific tasks. According to their results, the petal helped the users better recognize the delay time and delay type between airports. Moreover, the accuracy of recognizing the longest delay type in an airport is higher when the difference of delay time between the longest delay type and the second longest delay type is larger. The researchers than used the same hypotheses for the thread design. The results again showed that the larger the difference between the longest and the second longest delays, the higher accuracy and shorter time for users to recognize longest delay of an airport. However, they found the colors used for different threads could confuse their participants, so they added numeric information in the legend view that users can refer to.

Zhang et al. (2014) proposed a real-time visual analytics process based on microblog and emergency call data. The motivation of this research was to solve a challenge in the IEEE Symposium on Visual Analytics Science and Technology (VAST) 2014. The researchers built an extension for their previous work, SMART (Social Media Analytics and Reporting Toolkit) system (Cui, Chae, & Ebert, 2014). They proposed an integrated visual analytics framework to integrate different data sources and data attributes for event identification. Their system was also able to monitor the anomaly events by using microblog and emergency call stream. In this challenge, the researchers successfully identified several major events happened in Abila city. They also monitored the time lines and investigated the underlying connections. No explicit experimental digits were provided in this study.

2.3 Hashtag analysis

Feng et al. (2015) proposed STREAMCUBE, which focused on hierarchical spatio-temporal hashtag clustering techniques and generated hashtag clusters for automatically identifying potential events. In order to scale the large amount of Twitter information in different time frames and different areas, the researchers considered both space and time granularity in its database. STREAMCUBE was extended from the traditional data cube. They designed a single-pass clustering algorithm for event identification as well as a event ranking method to find burst events in real time. In their experiments, the researchers first compared the clustering quality of STREAMCUBE with other tweet clustering approaches. STREAMCUBE outperformed other algorithms by achieving 38.1% Normalized Mutual Information (NMI) and 71.7% Rand Index (RI). Second, they compared the performance of event ranking with other approaches. STREAMCUBE outperformed other algorithms by achieving 63.4% Mean Average Precision (MAP). Finally, the researchers also provided discussions about the scalability and the memory usage of their implementation of STREAMCUBE.

Anusha and Singh (2015) analyzed Twitter data based on the trending hashtags. The researchers attempted to find events that might intrigued specific users based on the hashtags used by the user and his sentimental states derived from the user's tweets. To analyze the tweets, the proposed system contained a topic modeling module to find the score of interestingness and a sentiment analysis module to detect the polarity. The researchers first extracted tweets by using specific hashtags as keywords. In topic modeling, they adopted Latent Dirichlet Allocation (LDA) to infer latent topics to which the tweets they collected had belonged. After getting the latent topics, they defined a tweet scoring measure to compute an interestingness score for every tweet. In the sentiment polarity analysis, they used NLTK corpora as training data and the NLTK analysis to decide the sentiment polarity of the tweets. In the experiments, the researchers provided a case study to discuss the interestingness scores and sentiment polarity (positive/negative) of the tweets related to hashtags about ICC World Cup 2015. However, since the number of topics considered for their dataset was low, they only showed several tweets with their positive/negative polarities and interestingness scores. No solid experimental results were provided in this study.

Wang and Iwaihara (2015) proposed a method to induce senses of a hashtag in a particular time frame on Twitter. The researchers built a co-occurrence graph that modeled the words and hashtags as nodes and the consecutiveness between two hashtags as an edge. They proposed a hashtag sense induction system aimed to extract a list of words with high node degree and used them to represent a sense of a community. For each hashtag, the system built a list of words as the induced senses of the hashtag. In their implementation, they took the entries in Wikipedia disambiguation list as Wikipedia senses. In the experiments, for each sense of a hashtag, the researchers used the context words extracted from the co-occurrence graph as keywords to fetch tweets, then ranked every tweet by the score that calculated by summing up all the weights of keywords existing in it. They asked human subjects to judge whether the sense of the hashtag in the top K tweets matches the sense they induced. The highest average precision of the system achieved 81.66% (K=10).

Pervin et al. (2015) performed an analysis on the co-occurrence of hashtags. The researchers designed the hypotheses to determine if the popularity of a hashtag increases when it appears along with one or more other similar hashtags. In order to evaluate the hashtag popularity, they defined several popularity factors (e.g. hashtag specific variables, dyad specific variables and control variables) for a hashtag. The researchers than perform experiments to determine if the scores of popularity factors are higher when a hashtag had other co-occurred ones. Their findings show that the popularity of a hashtag increases when a hashtag appears with other hashtags on the Great Eastern Japan earthquake tweet dataset. However, the criteria that they claim their hypotheses are true were not included in their study.

Cepni and Akan (2014) proposed a social sensing model for event detection and estimation with Twitter. The researchers modeled the information propagation on Twitter as a sensor network and adopted the communication theories to solve this problem. They considered an event as a sensor which with signal strength. The event estimation is made when an event signal is strong enough at a specific time (i.e., when relevant tweets accumulated in hashtag timeline.) The accuracy of the estimated signal is explored with mean square error analysis. In the experiment, the researchers calculate Mean Squared Error (MSE) in the estimation of every signal for various cases. The first experiment was designed by varying tweeting probabilities, hashtag use and network blockage probabilities, and concluded that tweeting probabilities impacted a lot more than hashtag use and network blockage probabilities. They also experimented on the number of active sensors and concluded that increasing the number of active sensors ultimately improved MSE. In the second experiment, they experimented on the effect of geotag use and suggested that the accuracy of information aggregated hashtags timeline improved 7-10% while users use geotag.

Denton et al. (2015) employed user hashtags to capture the description of image content of Facebook users. The researchers utilized the metadata, such as age, gender, home city and country of Facebook users combined with image features extracted from a convolutional neural network algorithm to predict the possible hashtags for images. They proposed three different models (i.e. bilinear model, user-based bilinear model and user-multiplicative tensor model) and trained those models by minimizing the weighted approximate-rank pairwise (WARP) loss. They evaluated their models respectively on 20 million public images uploaded on Facebook over a period of several days by using precision, recall and accuracy as the metrics. The user-multiplicative model outperformed other models. Their findings demonstrate how the user metadata combined with image features could be used for image hashtag prediction. Moreover, their user-multiplicative model gained the most significant performance boost.

Wang and Zheng (2014) performed a study of hashtag diffusion in Twitter. The researchers analyzed the hashtag diffusion by macro and micro perspectives. Their dataset contained 12 million tweets, but they kept only 153 hashtags that appeared in more than 100 tweets. From the macro perspective, they studied the diffusion by the tweet/hashtag properties combined with three proposed hashtag classes “single spike”, “multi-spikes” and “fluctuation”. They manually classified the 153 hashtags into three proposed hashtag classes and discussed the classes’ difference in terms of the general properties (e.g. hashtag length, tweet size, retweet ratio, etc.). From the micro perspective, the researchers adopted Edelman’s topology of influence (TOI) theory to discuss the individual diffusion. They characterized the tweet users as idea starter (IS.), amplifiers (Amp.), adapters (Ad.) and commentators (Com.) and defined four scores respectively according to TOI theory. They computed the four scores and discussed the role differences for the proposed hashtag classes. The researchers found that the user group that used the hashtags with the fluctuation pattern was formed by fewer idea starters and more commentators, which showed idea starters influenced others more in event-specific topics but less in general-interest topics. Second, the researchers studied properties of the users within different role categories. They used five network structure properties based on the user mention network, including in-degree, out-degree, total-degree, betweenness and in-out degree ratio, boundary spanner and the average number of tweets. According to the results, the researchers found that the idea starters have higher in-degree and the most uneven in-out degree ratio, which means they might be popular users who were mentioned by many others. The betweenness of the idea starters is also high because other users tend to re-share the contents posted by them. Finally, they also found that the adapters are the most active, which indicated the adapters are busier than other users because they combine and re-share ideas from many other users.

2.4 Summary

Among the above articles, STREAMCUBE (Feng et al., 2015) was the only study that mainly focused on tweet clustering based on hashtag to the best of our knowledge. STREAMCUBE proposed a detailed approach for tweet preprocessing, feature extraction, and a single-pass hashtag clustering algorithm. We were interested in how STREAMCUBE would perform when using our Twitter dataset related to the Paris Attacks. Also, we wanted to adopt a different hashtag clustering algorithm, such as K-means and see if we could extend the original STREAMCUBE and improve its performance by using different clustering algorithm. Finally, in order to evaluate the clustering results, we adopted Purity (Zhao & Karypis, 2001) and normalized mutual information (NMI) (Strehl & Ghosh, 2003) as the performance measures in this study. The details were as described in the next chapter.

CHAPTER 3. METHODOLOGY

3.1 Introduction

This study aimed to compare the clusters generated by our approach and by the existing event-detection approaches. Figure 3.1 is the workflow of this research. First, we collected tweets through Twitter public API and preprocess the data into features. Second, we implemented the K-means clustering algorithm as the clustering module as well as the clustering algorithm of STREAMCUBE (Feng et al., 2015) as another clustering module. Finally, we performed experiments and discussed the performance of the compared clustering approaches.

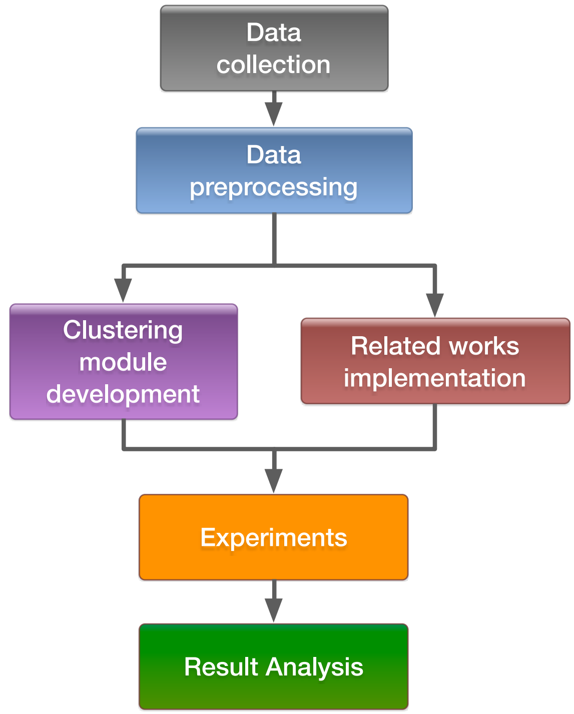

Figure 3.1. Workflow overview

Figure 3.2 depicts the system architecture of the proposed event detection method based on hashtag. We discuss our data collection and preprocessing for Twitter data in section 3.2. In section 3.3, we described the implementation details of the clustering modules of the K-means approach and the STREAMCUBE approach respectively. The metrics of performance evaluation are described in section 3.4.

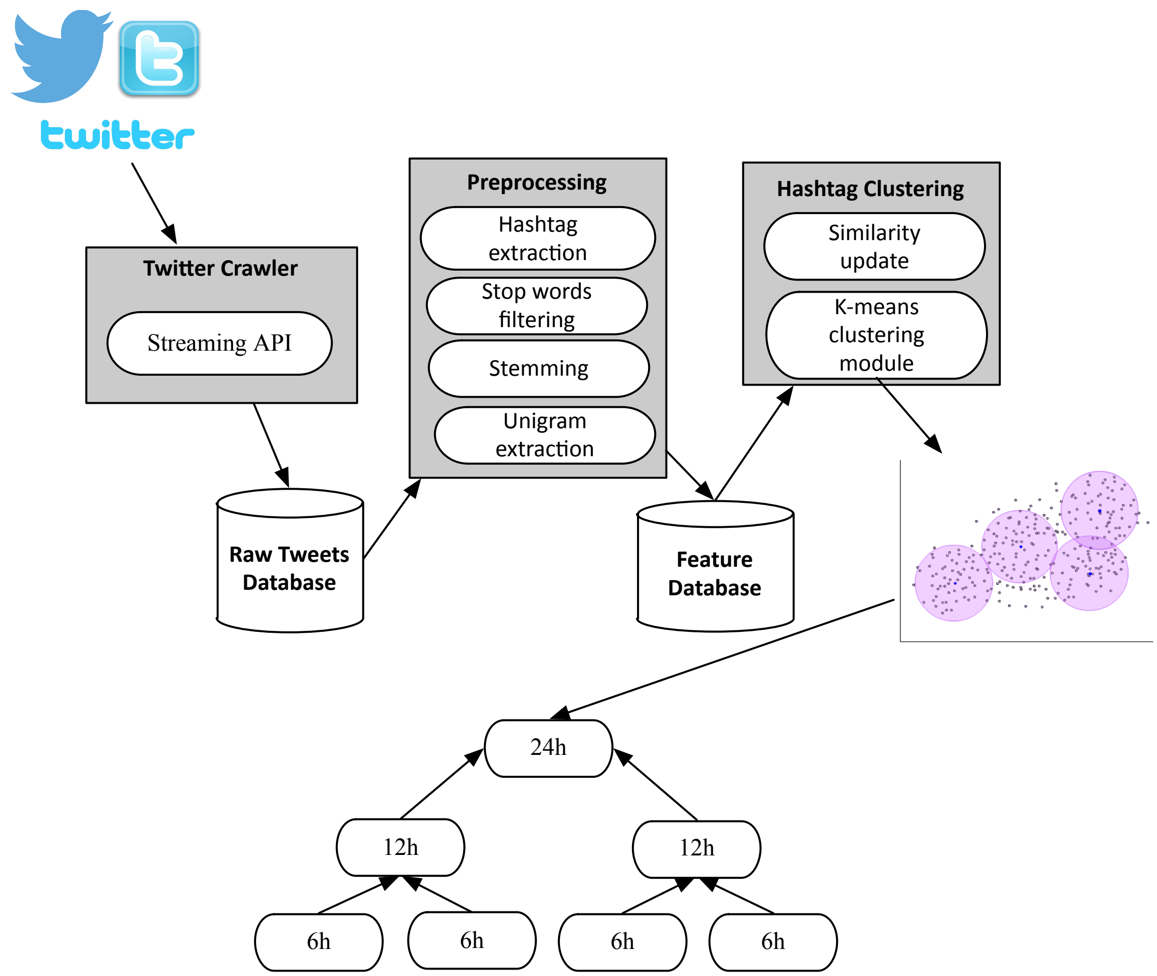

Figure 3.2. System architecture

3.2 Data collection and preprocessing

Twitter provided a set of streaming APIs that gave developers low latency access to its global stream of tweets. In this study, we used the Tweepy APIs (https://github.com/tweepy/tweepy), which was a Python library for accessing the Twitter. The APIs enabled us to collect the tweets related to a specific keyword list.

For each tweet, the following properties were collected: created time, number of retweet, text content, mentioned hyperlinks, mentioned hashtags and geographic coordinates. The preprocessing steps in figure 3.3 were used to extract features from the collected tweets.

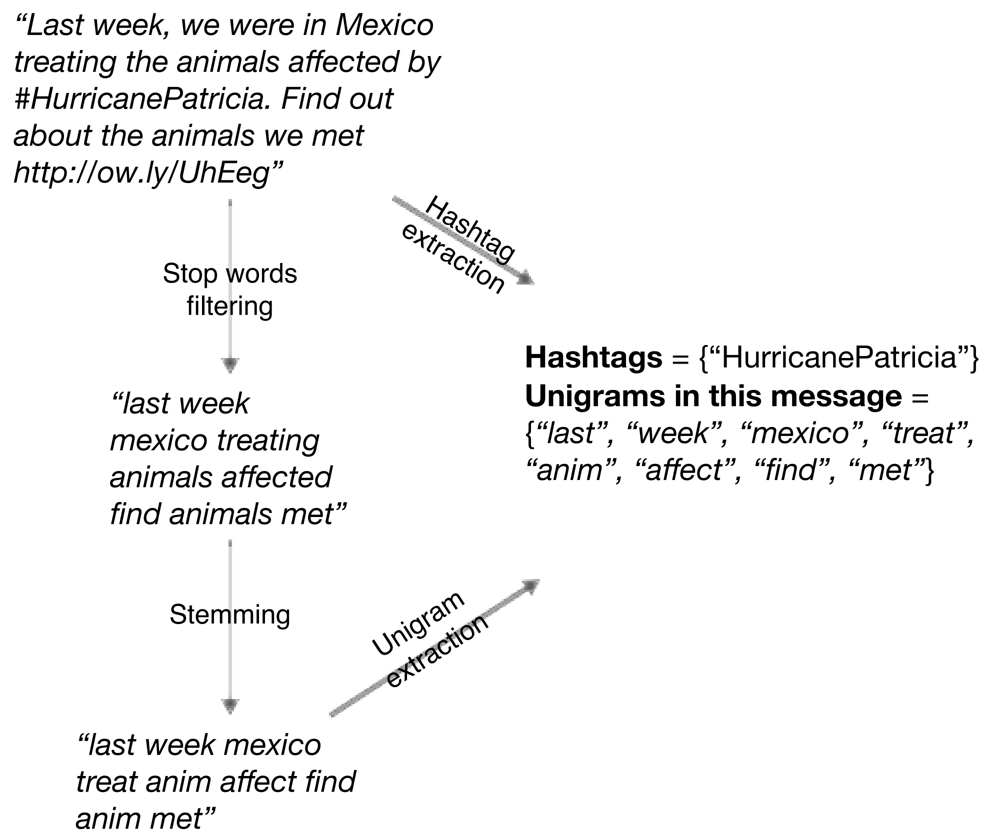

Figure 3.3. Tweet preprocessing

1. Hashtags in tweets were extracted as unigram features and removed from the original messages.
2. Lowercased all characters in tweets. Removed special characters, stop words, and hyperlinks.

3. All the tweets were stemmed using the Porter stemmer for reducing inflected words to their word stem.

The following example demonstrates how we extracted the features. Suppose that we have two tweet messages:

message1 = "Last week, we were in Mexico treating the animals affected by #HurricanePatricia. Find out about the animals we met http://ow.ly/UhEeg"

message2 = "The #Tornado last week was horrible!"

First, the hashtags are extracted and removed from the messages:

message1 = "Last week, we were in Mexico treating the animals affected by. Find out about the animals we met http://ow.ly/UhEeg", *hashtag* = {"HurricanePatricia"}

message2 = "The last week was horrible!", *hashtag* = {"Tornado"}

Second, lowercase all characters in tweets can then remove special characters, stop words, and hyperlinks:

message1 = "last week mexico treating animals affected find animals met", *hashtag* = {"HurricanePatricia"}

message2 = "last week horrible", *hashtag* = {"Tornado"}

Finally, after processed by Porter Stemmer, the results are:

message1 = "last week mexico treat anim affect find anim met", *hashtag* = {"HurricanePatricia"}

message2 = "last week horrible", *hashtag* = {"Tornado"}

3.3 Hashtag clustering

In the research, we adopted the feature definition of the hashtag clustering approach in STREAMCUBE. In section 3.3.1, we define the representations of hashtag and event. In section 3.3.2, we introduce how we adopt K-means cluster algorithm in this study.

3.3.1 Hashtag and event representations

First, a hashtag h can be considered as a bag of words, which is an aggregation of all the tweets that contain h. Let W denote all the words in our tweets, a hashtag h can be represented as a normalized weighted vector

$$\mathrm{h}_{tweet} = (w_1, w_2, \ldots, w_{|W|})$$

where w_i is the weight of the i-th word and $\|\, \mathrm{h}_{tweet} \,\| = 1$.

Second, a hashtag h can also be considered as a bag of hashtags because many hashtags have co-occurred with other hashtags. Let H denote the hashtag set, a hashtag h can be represented as a normalized weighted vector

$$\mathrm{h}_{tag} = (h_1, h_2, \ldots, h_{|H|})$$

where h_i is the weight of the i-th word and $\|\, \mathrm{h}_{tag} \,\| = 1$.

By using the above two representation, the distance between two hashtags can be defined. Let h^i_{tweet} and h^i_{tag} denote the word vector and hashtag vector of the i-th hashtag h_i . Given two hashtag h_1 and h_2, the distance is defined as

$$sim(\mathrm{h}_1, \mathrm{h}_2) = \alpha * cos(\mathrm{h}^1_{tweet}, \mathrm{h}^2_{tweet}) + \beta * cos(\mathrm{h}^1_{tag}, \mathrm{h}^2_{tag})$$

$$= (\alpha^{1/2}\, \mathrm{h}^1_{tweet}, \beta^{\,1/2}\, \mathrm{h}^1_{tag})\, (\alpha^{1/2}\, \mathrm{h}^2_{tweet}, \beta^{\,1/2}\, \mathrm{h}^2_{tag})$$

where α and β are two hyperparameters and $\alpha + \beta = 1$. From the above equation, a hashtag can be represented as a vector:

$$\mathrm{h} = (\alpha^{1/2}\, \mathrm{h}_{tweet}, \beta^{\,1/2}\, \mathrm{h}_{tag})$$

The following is a running example. We continue from the messages preprocessed before:

message1 = "last week mexico treat anim affect find anim met"

message2 = "last week horrible"

hashtag = {"HurricanePatricia", "Tornado"}

In the messages, W = {'last', 'week', 'mexico', 'treat', 'anim', 'affect', 'find', 'met', 'horrible'}, H = {"HurricanePatricia", "Tornado"}. The features of the hashtags are extracted as below:

$$\mathrm{h}_{\#HurricanePatricia} = (\alpha^{1/2}\, \mathrm{h}^{\#HurricanePatricia}{}_{tweet}, \beta^{\,1/2}\, \mathrm{h}^{\#HurricanePatricia}{}_{tag})$$

$$\mathrm{h}^{\#HurricanePatricia}{}_{tweet} = (0.11, 0.11, 0.11, 0.11, 0.22, 0.11, 0.11, 0.11, 0)$$

$$\mathrm{h}^{\#HurricanePatricia}{}_{tag} = (1, 0)$$

$$\mathrm{h}_{\#Tornado} = (\alpha^{1/2}\, \mathrm{h}^{\#Tornado}{}_{word}, \beta^{\,1/2}\, \mathrm{h}^{\#Tornado}{}_{tag})$$

$$\mathrm{h}^{\#Tornado}{}_{tweet} = (0.33, 0.33, 0, 0, 0, 0, 0, 0, 0.33)$$

$$\mathrm{h}^{\#Tornado}{}_{tag} = (0, 1)$$

Given that $\alpha = 0.5$ and $\beta = 0.5$, the distance between the hashtag "#HurricanePatricia" and "#Tornado" based *message1* and *message2* is:

$sim(h_{\#HurricanePatricia}, h_{\#Tornado})$

$= \alpha * cos(h^{\#HurricanePatricia}{}_{tweet}, h^{\#Tornado}{}_{tweet}) + \beta * cos(h^{\#HurricanePatricia}{}_{tag},$

$h^{\#Tornado}{}_{tag})$

$= 0.826$

3.3.2 K-means clustering

In the clustering algorithm of STREAMCUBE, the researchers designed a single-pass clustering algorithm because they aimed to process data in real time without using an iteration-based algorithm. This issue did not occur in this study because we concentrated on discovering the similarity and dissimilarity between different clustering methods but not the real time capability.

We adopted the K-means clustering method to find hashtag clusters by using the features described in section 3.3.1. We chose K-means because of the following reasons:

1. To identify the clusters, hierarchical clustering and K-means are two well-known cluster algorithms (Zhang et al., 2012). However, considering the great number of features used in our large-scale dataset, K-means is relatively faster and more effective.
2. Scaling K-means to massive data is relatively easy with respect to the algorithm's simplicity and iterative nature (Bahmani et al., 2012).

To perform K-means, we need the following parameters:

1. The distance function used to compute the distance between two points and the means of cluster centers. In this study, we use the distance function introduced in 3.3.1 as the distance function.
2. The selection of the number of clusters. In our experiment, we explored how to adequately set *K* for our dataset to gain the best performance. Since there is no perfect mathematical criterion exists (Jain, 2010), we experimented on how to set a best range of K values that could lead to better performance in the K-means approach.

In order to implement the clustering method, we adopted the K-means function in Natural Language Toolkit (Bird, 2006), a leading platform for building Python programs to work with human language data. The K-means toolkit provided the flexibility for programmers to use their own distance function instead of Euclidean distance. We used the following distance function introduced in 3.3.1 for two hashtags h_1, h_2.

$$sim(h_1, h_2) = \alpha * cos(h^1_{tweet}, h^2_{tweet}) + \beta * cos(h^1_{tag}, h^2_{tag})$$

$$= (\alpha^{1/2} h^1_{tweet}, \beta^{1/2} h^1_{tag}) (\alpha^{1/2} h^2_{tweet}, \beta^{1/2} h^2_{tag})$$

h^i_{tweet} and h^i_{tag} denote the word vector and hashtag vector of the *i*-th hashtag h_i. We set $\alpha = \beta = 0.5$. Every hashtag was represented as the following vector:

$$h = (\alpha^{1/2} h_{tweet}, \beta^{1/2} h_{tag})$$

We also used a similar feature vector to represent a cluster e:

$$e = (\alpha^{1/2} e_{tweet}, \beta^{1/2} e_{tag})$$

where e_{tweet} represents the average feature vector of every hashtag's tweet vectors in this cluster, and e_{tag} represents the average feature vector of every hashtag's h_{tag} vectors in this cluster. By defining the above functions, we were able to compute the distance between two hashtag, the distance between two cluster, as well as the distance between a cluster and a hashtag.

3.3.3 STREAMCUBE clustering

In order to compare with STREAMCUBE, we performed a rough implementation of their clustering approach as we eventually could not access the programs and datasets were not provided by the authors. Since the implementation was not identical to the original programs used in their experiments, the performance differences may have occurred in our implementation.

In STREAMCUBE, Feng et al. designed a single-pass hashtag clustering algorithm shown in figure 3.4. For each new hashtag, the algorithm first used a nearest-neighbor (shown in figure 3.5) function to find the existing cluster nearest to the hashtag. Second, the algorithm checked the absorbing condition to decide if the hashtag should be absorbed into the nearest cluster. If the distance between the hashtag and the nearest cluster was greater than the cluster's minimum threshold (i.e. the nearest distance between the cluster and any other clusters), the hashtag initialized a new cluster; Otherwise, the hashtag was absorbed by the cluster.

Algorithm 2: HASHTAG-CLUSTER-STATIC(E, h)

Input: Event set E={e_1, e_2, ..., e_k}
Hashtag h
Output: Updated event set E
e = nearest-neighbor(E, h)
if *sim(e, h) > e.threshold* **then**
  add h to E as a new event
else
  add h to the existing event e

Figure 3.4. HASHTAG-CLUSTER-STATIC algorithm (Feng et al., 2015)

Algorithm 3: NEAREST-NEIGHBOR(E, h)

Input: Event set E={e_1, e_2, ..., e_k}
Hashtag h={w_1, w_2, ..., w_n}
Output: Nearest neighbor e
for *each word w_i order by max_partial(w_i)* **do**
  for *each event e in the invert list of w_i* **do**
    $sim(h, e) \mathrel{+}= w_i^h * w_i^e$
  Let e^1 denote the most similar event
  Let e^2 denote the second similar event
  if *sim(h, e^1) > sim(h, e^2) +* $\sum_{j=i+1}^{n} max\_partial(w_j^e) \cdot w_j^h$ **then**
    return e^1

Figure 3.5. NEAREST-NEIGHBOR algorithm (Feng et al., 2015)

3.4 Data analysis

This research aimed to analyze similarity and difference between the cluster results generated in section 3.3.2 and 3.3.3.

In terms of cluster analysis, there is no best measure for evaluating the cluster quality (Bhatnagar & Ahuja, 2010). However, a mix of internal and external quality criteria provides us a comprehensive view to evaluate the clustering approaches. Therefore, we adopted two widely used metrics: Purity (Zhao & Karypis, 2001) as an external criterion and normalized mutual information (NMI) (Strehl & Ghosh, 2003) as an internal criterion to evaluate the quality of the clustering results.

Purity is an external quality criterion and is used when classes in the data are known. It measures the extent that if the documents in a cluster are from primarily one specific class. Given there are *k* clusters formed by total *n* documents that each document was labeled by one of *I* classes. The Purity of *r*-th cluster S_r with size n_r is defined as the equation in figure 3.6:

$$P(S_r) = \frac{1}{n_r} max_i(n_r^i), \forall i \in 1, 2, ..., |I|$$

Figure 3.6. The equation of the Purity (1)

where n_r^i is the number of documents of the *i*-th class that were assigned to the *r*-th cluster. The overall Purity of the clustering solution is obtained as a weighted sum of the individual cluster purities and is given by the equation in figure 3.7:

$$Purity = \sum_{r=1}^{k} \frac{n_r}{n} P(S_r)$$

Figure 3.7. The equation of the Purity (2)

In general, the larger the values of Purity, the better the clustering solution is.

NMI is an internal quality criterion and captures the commonality between two clustering approaches. It provides an indication of the shared information between a pair of clusters. Given X and Y be the random variables described by the cluster labeling $\lambda^{(a)}$ and $\lambda^{(b)}$, with $k^{(a)}$ and $k^{(b)}$ respectively. Let $I(X,Y)$ denote the mutual information between X and Y, $H(X)$, $H(Y)$ denote the entropy of X, Y. The equation of NMI is as the equation in figure 3.8:

$$NMI(X,Y) = \frac{I(X,Y)}{\sqrt{H(X)H(Y)}}$$

Figure 3.8. The equation of NMI (1)

Let $n_h^{(a)}$ be the number of objects in h-th cluster according to $\lambda^{(a)}$, $n_l^{(b)}$ be the number of objects in l-th cluster according to $\lambda^{(b)}$. Let $n_{h,l}$ denote the number of objects that are in h-th cluster according to $\lambda^{(a)}$ as well as in l-th cluster according to $\lambda^{(b)}$. Then the NMI $\phi^{(NMI)}$ can be rewritten as the following equation in figure 3.9:

$$\phi^{(\mathrm{NMI})}(\lambda^{(a)},\lambda^{(b)}) = \frac{\sum_{h=1}^{k^{(a)}}\sum_{\ell=1}^{k^{(b)}} n_{h,\ell}\log\left(\frac{n\cdot n_{h,\ell}}{n_h^{(a)} n_\ell^{(b)}}\right)}{\sqrt{\left(\sum_{h=1}^{k^{(a)}} n_h^{(a)}\log\frac{n_h^{(a)}}{n}\right)\left(\sum_{\ell=1}^{k^{(b)}} n_\ell^{(b)}\log\frac{n_\ell^{(b)}}{n}\right)}}$$

Figure 3.9. The equation of NMI (2)

The value of NMI is a fraction between 0 and 1, with 0 indicating that the two clusters do not shared the same information and 1 indicating that the two clusters are exactly the same.

CHAPTER 4. EXPERIMENTS

4.1. Data Collection

We collected 11,884,448 tweets during November 13, 2015 to November 17, 2015 for the Paris Attacks. The keyword list for collecting the tweets contained the following keywords: 'paris', 'attack', 'Gunmen', 'Bataclan', 'gunfire', 'hostage', 'Les Halles', 'Belle Equipe', 'Petite Cambodge', 'le Carillon'. We then filtered those tweets without text content, created time or geolocation to ensure the collected tweets did not lack any information we needed. Geolocation parameter is assumed to be useful for experiments beyond the scope of this thesis. There were 20,514 tweets with 8,616 different hashtags in our tweet collection after filtering those without geolocation. In the following sections, we describe the design of our experiments from three different perspectives: 1) Comparing K-means to STREAMCUBE, with STREAMCUBE as the ground truth, 2) comparing K-means to STREAMCUBE, with human serving as the ground truth, and 3) finding better K values for K-means, with human serving as the ground truth.

4.2. Compare K-means to STREAMCUBE (STREAMCUBE as the ground truth)

In this experiment, we compared the differences of clustering results between the K-means approach and STREAMCUBE with STREAMCUBE as the ground truth. In this experiment, we did not consider which approach was better but investigated the commonality between the two approaches. First, we followed the group setting of STREAMCUBE to group the collected tweets by their created time into 6-hours, 12-hours, and 24-hours groups respectively. The original reason for this setting is because STREAMCUBE only keeps events from the last six hours in memory for increment updates in their online system. The historical data are fixed and flushed into disk-based storage (Feng et al., 2015). Once every six-hours data go into disk-based storage, the system merges two six-hours data as a 12-hours data. The merge rule applied for the rest of the levels. Since their coarsest granularity is a day, the merge rule stops for 24-hours data. Although we did not aim to build a realtime system, we followed their setting to ensure the performance of STREAMCUBE was not influenced by a different group setting from its original. Second, we performed STREAMCUBE to cluster the tweet groups. Since STREAMCUBE generated a dynamic number of clusters for every tweet group, we recorded the numbers of clusters for all tweet groups in order to use the numbers as the K values in the K-means approach. Third, we performed the K-means approach to cluster the tweet groups. Finally, we took the clustering results of STREAMCUBE as the ground truth and the clustering results of the K-means approach as the predictions to calculate the NMI and Purity scores.

Table 4.1, 4.2 and 4.3 listed the Purity and NMI scores of every 6, 12, 24 hours respectively. In table 4.4, the Purity scores showed that over 70% of clusters generated by the K-means approach can be matched to corresponding clusters generated by STREAMCUBE, and the NMI scores showed the commonality between the results of the two clustering approaches are 57.8% for 24-hour groups, 69.6% for 12-hours groups, and 69.9% for 6-hours groups respectively. The K-means approach and STREAMCUBE did share a large portion of similar clustering results, but some significant performance differences are worth to be investigated. To further understand the differences, we designed the experiment in section 4.3 to use human labeled tweets for comparing the two approaches.

Table 4.1. The NMI and Purity scores of every 6 hours

Date	Hour range	Number of tweets	Number of clusters	Purity	NMI
2015/11/13	18:00 - 24:00	783	8	83.3%	74.8%
2015/11/14	0:00 - 6:00	1214	9	87.8%	78.8%
2015/11/14	6:00 - 12:00	1038	6	71.4%	51.3%
2015/11/14	12:00 - 18:00	1274	29	66.1%	79.0%
2015/11/14	18:00 - 24:00	1848	11	75.3%	59.7%
2015/11/15	0:00 - 6:00	1262	5	73.8%	63.6%
2015/11/15	6:00 - 12:00	1451	13	77.5%	69.1%
2015/11/15	12:00 - 18:00	1645	25	68.4%	74.3%
2015/11/15	18:00 - 24:00	1302	6	73.2%	60.1%
2015/11/16	0:00 - 6:00	1275	4	75.5%	63.7%
2015/11/16	6:00 - 12:00	1598	25	67.2%	74.3%
2015/11/16	12:00 - 18:00	1718	41	68.8%	80.3%
2015/11/16	18:00 - 24:00	808	11	84.1%	76.1%
2015/11/17	0:00 - 6:00	855	5	87.5%	83.2%
2015/11/17	6:00 - 12:00	431	5	88.9%	79.9%
2015/11/17	12:00 - 18:00	1191	6	63.8%	69.6%
2015/11/17	18:00 - 24:00	821	5	63.4%	51.1%

Table 4.2. The NMI and Purity scores of every 12 hours

Date	Hour range	Number of tweets	Number of clusters	Purity	NMI
2015/11/13	12:00 - 24:00	783	8	83.3%	74.8%
2015/11/14	0:00 - 12:00	2252	27	75.0%	76.8%
2015/11/14	12:00 - 24:00	3122	17	72.0%	61.8%
2015/11/15	0:00 - 12:00	2713	11	64.1%	49.5%
2015/11/15	12:00 - 24:00	2947	22	70.9%	65.6%
2015/11/16	0:00 - 12:00	2873	40	72.6%	79.4%
2015/11/16	12:00 - 24:00	2526	41	65.4%	73.0%
2015/11/17	0:00 - 12:00	1286	8	80.9%	68.5%
2015/11/17	12:00 - 24:00	2012	22	72.1%	77.0%

Table 4.3. The NMI and Purity scores of every 24 hours

Date	Number of tweets	Number of clusters	Purity	NMI
2015/11/13	783	8	83.3%	74.8%
2015/11/14	5374	26	67.6%	56.2%
2015/11/15	5660	2	99.3%	1.0%
2015/11/16	5399	84	68.5%	78.5%
2015/11/17	3298	46	67.1%	78.3%

Table 4.4. The summary of table 4.1, 4.2 and 4.3

Hour range	Avg. number of clusters	Avg. Purity	Avg. NMI
6	13.3	75.1%	69.9%
12	21.8	72.9%	69.6%
24	33.2	77.2%	57.8%

4.3. Compare K-means to STREAMCUBE (human serving as the ground truth)

We further compared the performance between the K-means approach and STREAMCUBE by using human labeled tweets as the ground truth. First, we randomly selected 200 from 3298 tweets, which contains 170 hashtags and 7185 unigrams, collected on November 17th, 2015. Second, we asked a human subject (a graduate student) to manually label categories for each of the 200 tweets. We instructed the subject to choose any text he wanted to label the tweets, but to use only one label for each tweet. The subject used six different labels in the labeling task: “Travel”, “Terrorism”, “Pray”, “Life”, “Hiring”, and “Others”. The label distribution of the 200 tweets is shown in figure 4.1. Third, we used similar steps in section 4.2 to performed clustering on the tweets collected on November 17th, 2015. We generated the clusters of STREAMCUBE and the clusters of the K-means approach respectively. Fourth, for each of the two cluster sets, we extracted the 200 labeled tweets and kept the cluster information of them. Thus, we had the clustering results of the 200 labeled tweets generated by the two approaches respectively, and we had the human labeled information of the 200 tweets as the ground truth. Finally, we calculated the NMI and Purity scores for the K-means approach and

STREAMCUBE respectively. Table 4.5 and 4.6 were the hashtags of the top 20 large clusters generated by the K-means approach and by STREAMCUBE. Table 4.7 was the performance comparison between the K-means approach and STREAMCUBE. We have shown that the K-means approach performed better than STREAMCUBE on both the Purity and NMI scores given the same number of clusters.

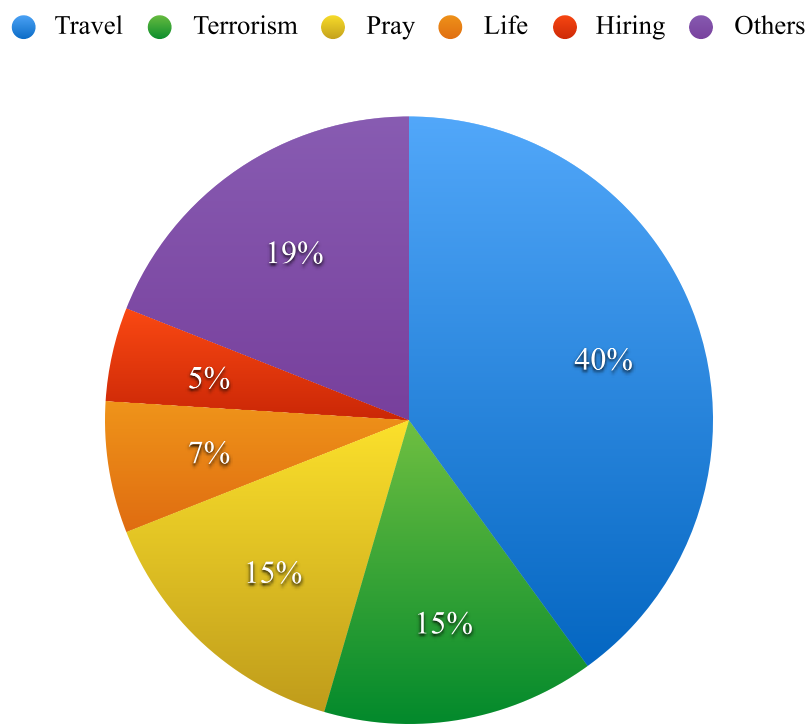

Figure 4.1. The distribution of the human labeled tweets

Table 4.5. Hashtags of the top 15 large clusters generated by the K-means approach

	K-means
1	parisattacks, igersparis, french, love, france, toulouse, city, picoftheday, photooftheday, pray, disneyland, fluctuatnecmergitur, europe, jesuisparis, peace, charliehebdo, parisian, tbt, view, prayforparis
2	SONIC, CareerArc, Retail, Lebanon, Job, job, Veterans, ExpediaJobs, LEBANON, CustomerService, Jobs, Sales, Hiring, Hospitality
3	Stigmabase, peaceforparis, informatique, vscocam, tb, movie, hope, instagood, stage, vsco, friends
4	bomb, ParisAttacks, parismaville, Adidas, portrait, COP21, jesuisenterrasse, quiz, Montemartre
5	TourEiffel, WeLoveParis, EiffelTower, MisterJoeCity, ILoveParis, Montmartre, France, Paris18, DirectLive
6	2DaysTilIKWYDLS, StreamMadeInTheAM, PrayForSyria, MTVStars, playpurpose, adtechNZ, SiyaKeRam, maritime
7	blue, frenchlife, me, parisstreet, iloveparis, ootd, metro, parisjetaime
8	tousaubistrot, concorde, attentat, parisattack, hommage, republique, placedelarepublique
9	selfies, streetlife, selfiewithart, streetart, photography, parisnights
10	london, football, huaweishot, Wembley, huawei, wembley
11	chezmatante, basket, creditmunicipal, inParis, villelumiere
12	like4like, toureiffel, beautiful, eiffeltower, bleublancrouge
13	christmas, beirut, lebanon, Beirut
14	news, London, hnytwtr, Fashion
15	PARIS, hiring, IT, Transportation
16	travel, vegas, JeSuisParis
17	bataclan, homage, rip
18	paris, Francia, prayfortheworld
19	burjkhalifa, mydubai, dubai
20	life, freedom

Table 4.6. Hashtags of the top 15 large clusters generated by STREAMCUBE

	STREAMCUBE
1	JeSuisParis, travel, tousaubistrot, bomb, news, parisattacks, Stigmabase, vegas, pray, fluctuatnecmergitur, eiffeltower, jesuisparis, peace, prayforparis, PrayForParis
2	PARIS, SONIC, CareerArc, Retail, Lebanon, hiring, Job, job, IT, Veterans, LEBANON, Transportation, Jobs, Hospitality
3	selfies, frenchie, selfie, parisstreet, streetlife, iloveparis, selfiewithart, frenchart, streetart, photography, parisnights, parisjetaime
4	2DaysTilIKWYDLS, trndnl, StreamMadeInTheAM, PrayForSyria, MTVStars, playpurpose, adtechNZ, SiyaKeRam, maritime
5	foodporn, ISIS, London, Adidas, Syria, COP21, Bataclan, ParisAttacks, movie
6	expo, chezmatante, basket, creditmunicipal, inParis, art, villelumiere
7	london, football, hnytwtr, huaweishot, Wembley, huawei, wembley
8	liberté, blue, music, liberteegalitefraternite, shym, bercy, ootd
9	attentat, parisattack, hommage, republique, freedom, placedelarepublique
10	TourEiffel, WeLoveParis, EiffelTower, ILoveParis, France, DirectLive
11	paris, Francia, prayfortheworld, Paris, vivelafrance
12	paris7, topparisphoto, iphone6plus, toureiffel
13	burjkhalifa, support, mydubai, dubai
14	ExpediaJobs, CustomerService, Sales, Hiring
15	french, parisian, tbt, friends
16	life, concorde, love, france
17	Repost, igersparis, usa, europe
18	parigi, photo, francia, beautiful
19	vscocam, vsco, disneyland
20	MisterJoeCity, Montmartre, Paris18

Table 4.7. The performance comparison based on human-labeled categories

Clustering approach	Number of clusters	Purity	NMI
K-means	46	70.5%	35.6%
STREAMCUBE	46	67.1%	27.8%

4.4. Find better K values for K-means (human serving as the ground truth)

We wanted to find the best K values for the K-means approach. In previous experiments, although we have shown the K-means approach could outperform STREAMCUBE when using the same number of clusters, the K values of the K-means approach were chosen based on the results of STREAMCUBE. In this experiment, we performed experiments on the K-means approach and compared the performance between different K-means. We again used the 200 manually labeled tweets created in section 4.3 as the ground truth and the clustering results as the prediction. In table 4.8, we performed the experiments for different K values. In figure 4.2 and 4.3, we found that both the Purity and NMI scores were higher when the number of clusters is larger.

Although the clustering method is better when the Purity is greater, high Purity is easy to achieve when the number of clusters is large. Thus, we should not use Purity to trade off the quality of the clustering against the number of clusters (Manning et al., 2008). The NMI scores reach 36% and become stable when the number of clusters is greater than 20. Moreover, in table 4.7, the Purity and NMI of STREAMCUBE was 67.1% and 27.8% while the number of cluster was 46. Results of table 4.8 show that once the number of clusters is greater than 20, the K-means approach could perform better than STREAMCUBE on both the Purity and NMI scores. Thus, according to our experiments, the K value for the K-means approach could be set at least greater than one tenth of the number of hashtags to achieve the performance better than STREAMCUBE.

Table 4.8. The performance of the K-means approach with different Ks

Number of clusters	Purity	NMI
2	49.4%	15.2%
5	53.9%	23.8%
10	59.7%	31.1%
15	66.5%	36.6%
20	67.8%	36.6%
30	70.0%	37.0%
50	70.0%	35.4%
100	73.8%	35.6%
150	75.6%	36.2%
170	76.8%	37.0%

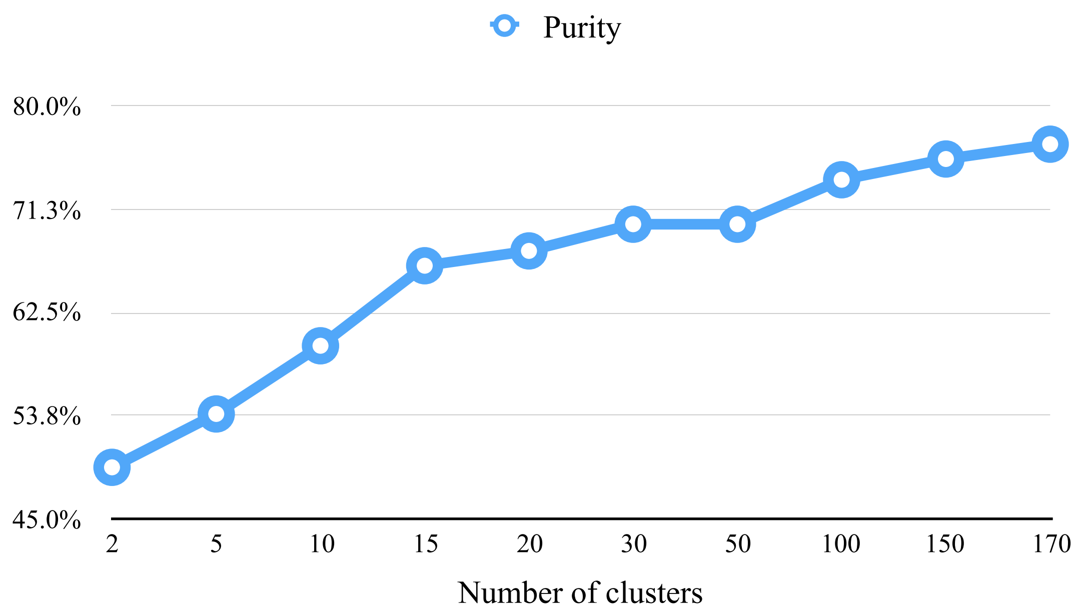

Figure 4.2. The line graph of Purity scores for different number of clusters

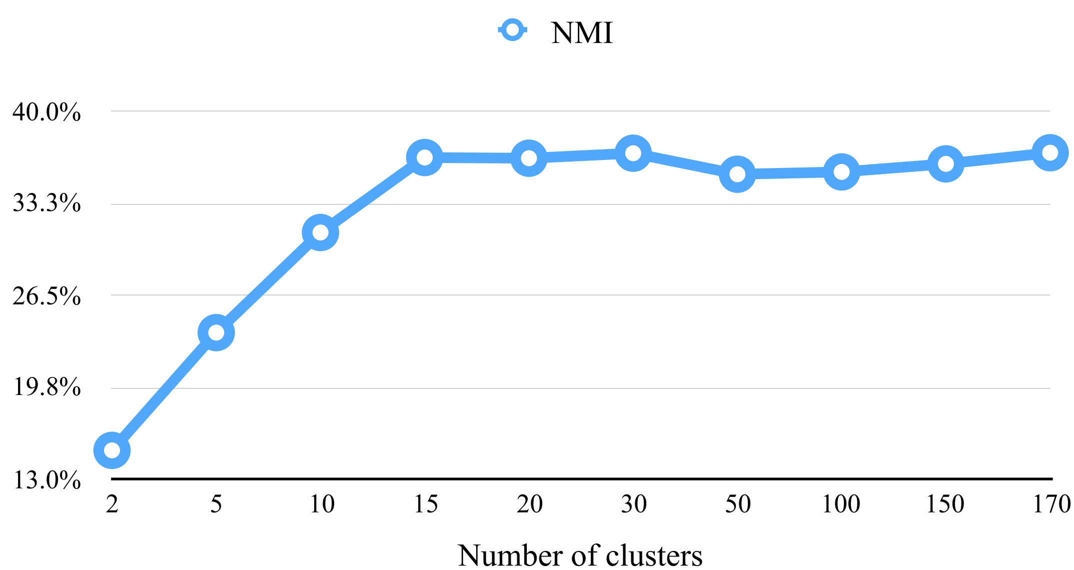

Figure 4.3. The line graph of NMI scores for different number of clusters

CHAPTER 5. CONCLUSION

5.1. Conclusion

In this study, we proposed an event detection approach that utilizes hashtags in tweets. We adopted the feature extraction used in STREAMCUBE and the clustering approach K-means. To the best of our knowledge, this is the first study to extend the framework of STREAMCUBE by adopting different clustering algorithm to enhance the original STREAMCUBE. Moreover, we collected the tweets related to the Paris Attack during November 13th to November 17th, 2015 as our datasets and performed the following experiments: First, we compared the commonality and difference between the K-means approach and STREAMCUBE in the perspectives of Purity and NMI on a full set of over 20,000 tweets. Second, we collected manual labels for 200 randomly sampled tweets from a human subject and demonstrated that the K-means approach outperformed STREAMCUBE on the clustering results. Third, we further discussed how to set the K value for the K-means approach to lead to a better clustering performance.

5.2. Recommendations for future studies

As with any study, some things could have been done better or more to further improve the results. The following are some recommendations for future research about this study.

1. Consider both spatial and temporal perspectives. In this study, we only considered the temporal perspective in the implementation and experiments. However, taking the spatial information into the implementation could make this study more practical for industry. For example, some events (e.g. local sport events) might not be discussed nationwide in lots of tweets but are densely discussed only in the tweets from certain areas. The characteristic could not be detected in our current approach. Considering both spatial and temporal perspectives could provide more information for event detection.
2. Compare with more related works of hashtag clustering. Our original idea was to compare our method with STREAMCUBE, Middleton et al. (2014), and Walther and Kaisser (2013). However, we only selected STREAMCUBE for some reasons: First, STREAMCUBE is the most relevant study to ours. Second, we could not find enough implementation details through the other two papers, and we were not able to reach the authors to use their original systems for our experiments. Although our approach was proved effective, a comprehensive comparison with the above different approaches could make our results more solid and convincing.

3. Collect more manual labels from human subjects for the collected tweets. In this study, we only collected manual labels for 200 tweets that were randomly sampled from the collected tweets of November 17th, 2015. Because the number of our manual labels and human subjects are low, inviting more subjects and increasing manually labeled tweets could enhance the credibility of our experiments when comparing our approach to others.
4. Analyze more events discussed on Twitter. We only collected the tweets related to the Paris Attack during November 13th to November 17th, 2015. Since this approach can be easily applied to any other events, it would be interesting if we could have applied our method on different datasets. If we could prove our approach also performs well on different datasets, it will demonstrate the practicality of this approach.
5. Adopt other clustering approaches in our hashtag clustering module instead of K-means. Although we have shown K-means is effective in our study and could perform better than the clustering approach of STREAMCUBE, the best setting of K values may be varied in different datasets and is not easy to be determined. Even though we have discovered how to set K values in this study, it is still difficult to transfer the knowledge to other datasets. It is worth to adopt different clustering approaches that could self-adapt to different datasets in our hashtag clustering module. Moreover, adopting multiple different clustering approaches will show the extensibility of this study.

REFERENCES

LIST OF REFERENCES

Anusha, A., & Singh, S. (2015, August). Is that twitter hashtag worth reading. In *Proceedings of the Third International Symposium on Women in Computing and Informatics* (pp. 272-277). ACM.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R., & Vassilvitskii, S. (2012). Scalable k-means++. *Proceedings of the VLDB Endowment,* 5(7), 622-633.

Bhatnagar, V., & Ahuja, S. (2010). Robust clustering using discriminant analysis. In *Advances in Data Mining. Applications and Theoretical Aspects* (pp. 143-157). Springer Berlin Heidelberg.

Bird, S. (2006, July). NLTK: The natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions* (pp. 69-72). Association for Computational Linguistics.

Cataldi, M., Di Caro, L., & Schifanella, C. (2010, July). Emerging topic detection on Twitter based on temporal and social terms evaluation. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining* (p. 4). ACM.

Cepni, K., & Akan, O. B. (2014, December). Social sensing model and analysis for event detection and estimation with Twitter. In *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2014 IEEE 19th International Workshop on* (pp. 31-35). IEEE.

Chiang, M. M. T., & Mirkin, B. (2010). Intelligent choice of the number of clusters in K-Means clustering: An experimental study with different cluster spreads. *Journal of classification, 27(1)*, 3-40.

Cui, Y., Chae, J., & Ebert, D. (2014). Social Media Analytics Reporting Toolkit.

Denton, E., Weston, J., Paluri, M., Bourdev, L., & Fergus, R. (2015, August). User Conditional Hashtag Prediction for Images. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1731-1740). ACM.

Feng, W., Han, J., Wang, J., Aggarwal, C., & Huang, J. (2015). STREAMCUBE: Hierarchical spatio-temporal hashtag clustering for event exploration over the Twitter stream. ICDE.

Hua, T., Chen, F., Zhao, L., Lu, C. T., & Ramakrishnan, N. (2013, August). STED: Semi-supervised targeted-interest event detectionin in Twitter. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1466-1469). ACM.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters, 31(8)*, 651-666.

Jung, K., Heo, W., & Chen, W. (2012, December). Irie: Scalable and robust influence maximization in social networks. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on* (pp. 918-923). IEEE.

Ko, S., Afzal, S., Walton, S., Yang, Y., Chae, J., Malik, A., ... & Ebert, D. (2014, October). Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (pp. 83-92). IEEE.

Lee, R., & Sumiya, K. (2010, November). Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks* (pp. 1-10). ACM.

Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C. C. (2012, April). Tedas: A Twitter-based event detection and analysis system. In *Data engineering (icde), 2012 ieee 28th international conference on* (pp. 1273-1276). IEEE.

Liu, Y., Xu, S., & Tourassi, G. (2015). Detecting rumors through modeling information propagation networks in a social media environment. In *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 121-130). DC, USA. Springer International Publishing.

Lloyd, S. P. (1982). Least squares quantization in PCM. *Information Theory, IEEE Transactions on, 28*(2), 129-137.

Manning, C. D., & Raghavan, P. H. Schutze. 2009. An introduction to information retrieval.

Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. (2011, May). Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 227-236). ACM.

Mathioudakis, M., & Koudas, N. (2010, June). Twittermonitor: trend detection over the Twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (pp. 1155-1158). ACM.

Middleton, S. E., Middleton, L., & Modafferi, S. (2014). Real-time crisis mapping of natural disasters using social media. *Intelligent Systems, IEEE, 29*(2), 9-17.

Pervin, N., Phan, T. Q., Datta, A., Takeda, H., & Toriumi, F. (2015). Hashtag popularity on Twitter: analyzing co-occurrence of multiple hashtags. In *Social Computing and Social Media* (pp. 169-182). Springer International Publishing.

Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103-119.

Ritter, A., Etzioni, O., & Clark, S. (2012, August). Open domain event extraction from Twitter. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1104-1112). ACM.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010, April). Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851-860). ACM.

StatSoft (2010). Finding the right number of clusters in K-means and EM clustering: V-fold cross-validation. Retrieved December 14, 2015, from http://www.statsoft.com/textbook/cluster-analysis/.

Strehl, A., & Ghosh, J. (2003). Cluster ensembles---a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research, 3*, 583-617.

Walther, M., & Kaisser, M. (2013). Geo-spatial event detection in the Twitter stream. In *Advances in Information Retrieval* (pp. 356-367). Berlin, Germany. Springer Berlin Heidelberg.

Wang, F., Wang, H., & Xu, K. (2012, June). Diffusive logistic model towards predicting information diffusion in online social networks. In *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on* (pp. 133-139). IEEE.

Wang, M., & Iwaihara, M. (2015). Hashtag sense induction based on co-occurrence graphs. In *Web Technologies and Applications* (pp. 154-165). Springer International Publishing.

Wang, Y., & Zheng, B. (2014, August). On macro and micro exploration of hashtag diffusion in Twitter. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on* (pp. 285-288). IEEE.

Yedla, M., Pathakota, S. R., & Srinivasa, T. M. (2010). Enhancing K-means clustering algorithm with improved initial center. *International Journal of computer science and information technologies*, 1(2), 121-125.

Zhang, J., Afzal, S., Breunig, D., Xia, J., Zhao, J., Sheeley, I., ... & Chen, Y. (2014, October). Real-time identification and monitoring of abnormal events based on microblog and emergency call data using SMART. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (pp. 393-394). IEEE.

Zhang, Y., Wu, Y., & Yang, Q. (2012). Community discovery in Twitter based on user interests. *Journal of Computational Information Systems*, 8(3), 991-1000.

Zhao, Y., & Karypis, G. (2001). *Criterion functions for document clustering: Experiments and analysis* (Vol. 1, p. 40). Technical report.

APPENDICES

APPENDIX A: NUMBER OF ELEMENTS IN THE CLUSTERS

1. The number of elements in clusters (every 6 hours)

Date	Hour range	Number of clusters	Number of hashtags	Average number of hashtags in clusters	Standard Deviation of number of hashtags in clusters
2015/11/13	18:00 - 24:00	8	24	3.0	3.3
2015/11/14	0:00 - 6:00	9	41	4.6	5.6
2015/11/14	6:00 - 12:00	6	35	5.8	5.8
2015/11/14	12:00 - 18:00	29	56	1.9	1.5
2015/11/14	18:00 - 24:00	11	89	8.1	8.9
2015/11/15	0:00 - 6:00	5	65	13.0	8.6
2015/11/15	6:00 - 12:00	13	71	5.5	5.9
2015/11/15	12:00 - 18:00	25	79	3.2	3.4
2015/11/15	18:00 - 24:00	6	71	11.8	8.4
2015/11/16	0:00 - 6:00	4	53	13.3	5.0
2015/11/16	6:00 - 12:00	25	64	2.6	2.9
2015/11/16	12:00 - 18:00	41	96	2.3	2.0

Date	Hour range	Number of clusters	Number of hashtags	Average number of hashtags in clusters	Standard Deviation of number of hashtags in clusters
2015/11/16	18:00 - 24:00	11	44	4.0	4.4
2015/11/17	0:00 - 6:00	5	32	6.4	3.5
2015/11/17	6:00 - 12:00	5	18	3.6	1.6
2015/11/17	12:00 - 18:00	6	58	9.7	5.3
2015/11/17	18:00 - 24:00	5	41	8.2	4.0

2. The number of elements in clusters (every 12 hours)

Date	Hour range	Number of clusters	Number of hashtags	Average number of hashtags in clusters	Standard Deviation of number of hashtags in clusters
2015/11/13	12:00 - 24:00	8	24	3.0	3.3
2015/11/14	0:00 - 12:00	27	80	3.0	4.6
2015/11/14	12:00 - 24:00	17	157	9.2	11.8
2015/11/15	0:00 - 12:00	11	131	11.9	16.7
2015/11/15	12:00 - 24:00	22	148	6.7	8.9
2015/11/16	0:00 - 12:00	40	113	2.8	3.8
2015/11/16	12:00 - 24:00	41	133	3.2	3.6
2015/11/17	0:00 - 12:00	8	47	5.9	4.4
2015/11/17	12:00 - 24:00	22	104	4.7	3.6

3. The number of elements in clusters (every 24 hours)

Date	Number of clusters	Number of hashtags	Average number of hashtags in clusters	Standard Deviation of number of hashtags in clusters
2015/11/13	8	24	3.0	3.3
2015/11/14	26	241	9.3	16.9
2015/11/15	2	303	151.5	149.5

Date	Number of clusters	Number of hashtags	Average number of hashtags in clusters	Standard Deviation of number of hashtags in clusters
2015/11/16	84	238	2.8	4.3
2015/11/17	46	170	3.7	3.4

APPENDIX B: TWEETS MANUALLY LABELED BY A HUMAN EXPERT

In this section, we listed the 200 tweets used in section 4.3 that were manually labeled by a human expert. For each tweet, the attribute “manual_label” contains a number between 1 to 6: 1 stands for “Travel”, 2 stands for “Terrorism”, 3 stands for “Pray”, 4 stands for “Life”, 5 stands for “Hiring”, and 6 stands for “Others”. The details of the tweets are as below:

1. {"original_text": "Barnie pense... \ud83d\ude1d\ud83d\ude3b #barnie @ Paris, France https://t.co/hxiXx8rCWa", "manual_label": "1", "hashtags": ["barnie"], "filtered_text": "barni pense ie paris franc wa", "original_id": 666681423290609664, "created_time": "Tue Nov 17 18:16:52 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}
2. {"original_text": "#LasVegas #Paris #Casino #whathappensinvegas #StayInVegas @ Paris Las Vegas Hotel & Casino https://t.co/ao9uAW7FNy", "manual_label": "1", "hashtags": ["LasVegas", "Paris", "Casino", "whathappensinvegas", "StayInVegas"], "filtered_text": "pari la vega hotel amp casino", "original_id": 666820776801263616, "created_time": "Wed Nov 18 03:30:36 +0000 2015", "geo": {"type": "Point", "coordinates": [36.11235778, -115.17147064]}}
3. {"original_text": "#paris #love #amour charlottebentz \ud83d\ude18 @ Paris Quai De Seine https://t.co/Ybrovm7K7o", "manual_label": "1", "hashtags": ["paris", "love", "amour"], "filtered_text": "charlottebentz pari quai de sein o", "original_id": 666552264937635840, "created_time": "Tue Nov 17 09:43:38 +0000 2015", "geo": {"type": "Point", "coordinates": [48.84867601, 2.35944041]}}
4. {"original_text": "Macken: Paris attacks \u2018wake up call\u2019 for more surveillance https://t.co/seUOxvCjFC #macken", "manual_label": "2", "hashtags": ["macken"], "filtered_text": "macken pari attack wake call surveil", "original_id": 666726781013983234, "created_time": "Tue Nov 17 21:17:06 +0000 2015", "geo": {"type": "Point", "coordinates": [64.7079076, 21.00379944]}}

5. {"original_text": "#vangogh #selfie #mentalillness #parisjetaime #paris #frenchart @ Mus\u00e9e d'Orsay (officiel) https://t.co/eHt6UPdwxA", "manual_label": "1", "hashtags": ["vangogh", "selfie", "mentalillness", "parisjetaime", "paris", "frenchart"], "filtered_text": "mus d orsay officiel", "original_id": 666808304442679297, "created_time": "Wed Nov 18 02:41:03 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86025591, 2.32602851]}}
6. {"original_text": "Good Morning Paris, Bonjo\u00far ! \ud83c\uddeb\ud83c\uddf7 #paris #france #prayforparis\ud83c\uddeb\ud83c\uddf7 @ A\u00e9roport de Paris-Charles-de-Gaulle\u2026 https://t.co/E4tdkzVU4f", "manual_label": "3", "hashtags": ["paris", "france", "prayforparis"], "filtered_text": "good morn paris bonjo r c is a roport de paris charles de gaul vu4f", "original_id": 666529925092917248, "created_time": "Tue Nov 17 08:14:52 +0000 2015", "geo": {"type": "Point", "coordinates": [49.01478417, 2.54166164]}}
7. {"original_text": "\ud83c\uddeb\ud83c\uddf7 #resistinpeace #prayforparis #republique #paris @ Place de la Republique https://t.co/mKtl48JRIQ", "manual_label": "3", "hashtags": ["resistinpeace", "prayforparis", "republique", "paris"], "filtered_text": "c u place de la republiqu iq", "original_id": 666741957859897345, "created_time": "Tue Nov 17 22:17:24 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86746503, 2.36418438]}}
8. {"original_text": "6. Syria\n7. #SiyaKeRam\n8. #APEC2015\n9. Paris\n10. B.A.P\n\n2015/11/17 14:15 SGT #trndnl https://t.co/psP0GzBgZB", "manual_label": "6", "hashtags": ["SiyaKeRam", "APEC2015", "trndnl"], "filtered_text": "6 syria 7 8 9 pari 10 b a p 2015 11 17 14 15 sgt", "original_id": 666501178952314880, "created_time": "Tue Nov 17 06:20:38 +0000 2015", "geo": {"type": "Point", "coordinates": [1.3656, 103.8277]}}
9. {"original_text": "ENSEMBLE \ud83d\udc97 \n\n#France #Paris #parisiloveyou #hope #crossfit #community #crossfitenyeto\u2026 https://t.co/HctNbtjyej", "manual_label": "4", "hashtags": ["France", "Paris", "parisiloveyou", "hope", "crossfit", "community", "crossfitenyeto"], "filtered_text": "ensembl e u e y o j", "original_id": 666547332901502976, "created_time": "Tue Nov 17 09:24:02 +0000 2015", "geo": {"type": "Point", "coordinates": [45.77353164, 4.79741492]}}
10. {"original_text": "England paid their respects to the victims of the Paris attacks during training ahead of their friendly tie with France #PrayForParis", "manual_label": "2", "hashtags": ["PrayForParis"], "filtered_text": "england paid respect victim pari attack train ahead friendli tie franc", "original_id": 666520695568998400, "created_time": "Tue Nov 17 07:38:11 +0000 2015", "geo": {"type": "Point", "coordinates": [8.83134, 3.74247]}}
11. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/AgQW5WpJ9E",

"manual_label": "4", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666801491659747328, "created_time": "Wed Nov 18 02:13:58 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

12. {"original_text": "Paris International Film Fantastic Festival ! #pifff #cinema #movie #film #fantastic #grandrex\u2026 https://t.co/LZzDQFiOSo", "manual_label": "1", "hashtags": ["pifff", "cinema", "movie", "film", "fantastic", "grandrex"], "filtered_text": "pari intern film fantast festiv", "original_id": 666686011896930304, "created_time": "Tue Nov 17 18:35:06 +0000 2015", "geo": {"type": "Point", "coordinates": [48.87050278, 2.347725]}}
13. {"original_text": "#Stigmabase | VN - Time for action, not words, from peace-loving Muslims in face of Paris barbarism \u00a0- But if you\u2026 https://t.co/tutDt52kXw", "manual_label": "2", "hashtags": ["Stigmabase"], "filtered_text": "vn time action words peace lov muslim face pari barbar but you", "original_id": 666771814043901952, "created_time": "Wed Nov 18 00:16:03 +0000 2015", "geo": {"type": "Point", "coordinates": [21.10748654, 105.85161133]}}
14. {"original_text": "#prayforparis #attentat #metz #paris #fabert \"nous devons rester fort...\"\u2026 https://t.co/fJY29j4yyL", "manual_label": "3", "hashtags": ["prayforparis", "attentat", "metz", "paris", "fabert"], "filtered_text": "nou devon rester fort", "original_id": 666592670626660353, "created_time": "Tue Nov 17 12:24:12 +0000 2015", "geo": {"type": "Point", "coordinates": [49.12199588, 6.17163429]}}
15. {"original_text": "Pas de panique ce ne sont que des nouilles de riz. #Wokbar (@ Wokbar in Paris, \u00cele-de-France) https://t.co/a27R9LfHtz", "manual_label": "1", "hashtags": ["Wokbar"], "filtered_text": "pa de paniqu ce ne sont que de nouill de riz wokbar paris le de france", "original_id": 666586903689502720, "created_time": "Tue Nov 17 12:01:17 +0000 2015", "geo": {"type": "Point", "coordinates": [48.84435579, 2.33124733]}}
16. {"original_text": "6. #SiyaKeRam\n7. #APEC2015\n8. Paris\n9. B.A.P\n10. LITTLE MIX\n\n2015/11/17 13:35 SGT #trndnl https://t.co/psP0GzBgZB", "manual_label": "6", "hashtags": ["SiyaKeRam", "APEC2015", "trndnl"], "filtered_text": "6 7 8 pari 9 b a p 10 littl mix 2015 11 17 13 35 sgt", "original_id": 666491119715291136, "created_time": "Tue Nov 17 05:40:40 +0000 2015", "geo": {"type": "Point", "coordinates": [1.3656, 103.8277]}}
17. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/I7GrS20woT", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari",

"original_id": 666804965160054785, "created_time": "Wed Nov 18 02:27:46 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

18. {"original_text": "#hnytwtr id53,IT,{2}WEB_X,Webmail attack from from 151.8.222.x req=|administrator|components|com_acymailing|inc|openflas...", "manual_label": "6", "hashtags": ["hnytwtr"], "filtered_text": "id53 it 2 web_x webmail attack 151 8 222 x req administrator components com_acymailing inc openflas", "original_id": 666570840130105344, "created_time": "Tue Nov 17 10:57:27 +0000 2015", "geo": {"type": "Point", "coordinates": [42.83, 12.83]}}

19. {"original_text": "@TCSITWiz i feel u are a global organization so u must change ur dp to show support for paris #tcsitwiz", "manual_label": "2", "hashtags": ["tcsitwiz"], "filtered_text": "feel u global organ u must chang ur dp show support pari", "original_id": 666496641180962816, "created_time": "Tue Nov 17 06:02:36 +0000 2015", "geo": {"type": "Point", "coordinates": [26.8706463, 80.9842061]}}

20. {"original_text": "My everything \ud83c\uddeb\ud83c\uddf7\u2764\ufe0f #sisters #sisterlove margauxf90 @pipa160 #paris #france #familyfirst #jeunesse\u2026 https://t.co/dtmQovygso", "manual_label": "4", "hashtags": ["sisters", "sisterlove", "paris", "france", "familyfirst", "jeunesse"], "filtered_text": "my everyth r ve margauxf9 6 c se", "original_id": 666714412040847361, "created_time": "Tue Nov 17 20:27:57 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86537748, 2.2736029]}}

21. {"original_text": "The tweet with the most impact of the 'Fabien Clain' Trend, was published by @mathieuvonrohr: https://t.co/hT9kgEXeGc (94 RTs) #trndnl", "manual_label": "2", "hashtags": ["trndnl"], "filtered_text": "the tweet impact fabien clain trend publish 94 rts", "original_id": 666598694859354112, "created_time": "Tue Nov 17 12:48:08 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8569, 2.3412]}}

22. {"original_text": "\u4f55\u3057\u306b\u6765\u305f\u306e\u304b\u5fd8\u308c\u308b\u3068\u3053\u308d\u3067\u3057\u305f\u3002\n\n\u30ec\u30b9\u30c8\u30e9\u30f3higuma negui ramen \n#paris #instagood #ramen #sakumanaohito @ Restaurant\u2026 https://t.co/IpYdG2ED9e", "manual_label": "1", "hashtags": ["paris", "instagood", "ramen", "sakumanaohito"], "filtered_text": "higuma negui ramen restaurant", "original_id": 666559248998883328, "created_time": "Tue Nov 17 10:11:23 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8634415, 2.33483]}}

23. {"original_text": "After Paris attacks, English football fans salute France by roaring out the ... -\u2026 https://t.co/TfTZutjDmc #news https://t.co/VzAvlZDhEl", "manual_label": "2", "hashtags": ["news"], "filtered_text": "after pari attacks english footbal fan salut franc roar https t co vzavlzdhel", "original_id":

666743709585666048, "created_time": "Tue Nov 17 22:24:22 +0000 2015", "geo": {"type": "Point", "coordinates": [36.06801502, -79.79721587]}}

24. {"original_text": "https://t.co/A88bO1P3BV #beirut", "manual_label": "6", "hashtags": ["beirut"], "filtered_text": "", "original_id": 666762934455017472, "created_time": "Tue Nov 17 23:40:46 +0000 2015", "geo": {"type": "Point", "coordinates": [33.89376063, 35.55259466]}}
25. {"original_text": "The #world #loves #Adidas #collectors #only : Swag #Adidas #Stripes #Bomb #Weapon #Army #Battlefield #Artwork #War", "manual_label": "2", "hashtags": ["world", "loves", "Adidas", "collectors", "only", "Adidas", "Stripes", "Bomb", "Weapon", "Army", "Battlefield", "Artwork", "War"], "filtered_text": "the swag", "original_id": 666771547944837121, "created_time": "Wed Nov 18 00:14:59 +0000 2015", "geo": {"type": "Point", "coordinates": [40.892601, -78.215897]}}
26. {"original_text": "6. Fran\u00e7a\n7. #PrayForSyria\n8. #2DaysTilIKWYDLS\n9. Jessica\n10. Paris\n\n2015/11/17 07:35 WET #trndnl https://t.co/uLzQlByvJf", "manual_label": "6", "hashtags": ["PrayForSyria", "2DaysTilIKWYDLS", "trndnl"], "filtered_text": "6 fran a 7 8 9 jessica 10 pari 2015 11 17 07 35 wet", "original_id": 666521272596099072, "created_time": "Tue Nov 17 07:40:29 +0000 2015", "geo": {"type": "Point", "coordinates": [38.9901, -9.1413]}}
27. {"original_text": "#PwC lighting up for #Paris @ PwC https://t.co/vM7k8AjO3r", "manual_label": "4", "hashtags": ["PwC", "Paris"], "filtered_text": "light pwc", "original_id": 666485906086469632, "created_time": "Tue Nov 17 05:19:57 +0000 2015", "geo": {"type": "Point", "coordinates": [37.3268623, -121.8890991]}}
28. {"original_text": "#UnitedWeStand #g\u00e9n\u00e9rationBataclan #PrayForTheWorld #WarIsComing #ParisAttacks #FrenchRepublic #Paris\u2026 https://t.co/XcmTqc2NdR", "manual_label": "2", "hashtags": ["UnitedWeStand", "g\u00e9n\u00e9rationBataclan", "PrayForTheWorld", "WarIsComing", "ParisAttacks", "FrenchRepublic", "Paris"], "filtered_text": "", "original_id": 666744393030893568, "created_time": "Tue Nov 17 22:27:05 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}
29. {"original_text": "This is a freedom place.\nWords are powerful.\nExpress yourself. \n\nLOVE , EMPATHY, FORGIVENESS\n\n#Paris @\u2026 https://t.co/ 9D7XVPXVCT", "manual_label": "3", "hashtags": ["Paris"], "filtered_text": "thi freedom place word powerful express yourself love empathy forgiv", "original_id": 666692648971833344, "created_time": "Tue Nov 17 19:01:28 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86746503, 2.36418438]}}
30. {"original_text": "Beautiful Washington Square Arch lit up on Sunday night in #NYC as a memorial for the Paris attacks\u2026 https://t.co/w8sNkEdWJQ", "manual_label": "3", "hashtags": ["NYC"], "filtered_text": "beauti washington squar arch lit sunday night memori pari attacks", "original_id": 666742524892061697,

"created_time": "Tue Nov 17 22:19:40 +0000 2015", "geo": {"type": "Point", "coordinates": [40.73083565, -73.9974153]}}

31. {"original_text": "L'#ob\u00e9lisque de la #concorde #paris #France @ Place de la concorde https://t.co/tTvZl9MGiI", "manual_label": "1", "hashtags": ["ob\u00e9lisque", "concorde", "paris", "France"], "filtered_text": "l de la place de la concord", "original_id": 666714627992981504, "created_time": "Tue Nov 17 20:28:48 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86562942, 2.32172938]}}

32. {"original_text": "#Montreal to #Paris with live #ParisAttacks @ Place \u00c9milie-Gamelin https://t.co/myQ1Gt6sHS", "manual_label": "1", "hashtags": ["Montreal", "Paris", "ParisAttacks"], "filtered_text": "live place milie gamelin", "original_id": 666819222836674560, "created_time": "Wed Nov 18 03:24:26 +0000 2015", "geo": {"type": "Point", "coordinates": [45.515414, -73.55996]}}

33. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/H0KqlhWGii", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666810152058621952, "created_time": "Wed Nov 18 02:48:23 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

34. {"original_text": "Paris attacks: All faiths need to stand together, says Bishop of #Shrewsbury https://t.co/sXO5xtoomE #Shropshire https://t.co/tjOeJtvZZl", "manual_label": "2", "hashtags": ["Shrewsbury", "Shropshire"], "filtered_text": "pari attacks all faith need stand together say bishop https t co tjoejtvzzl", "original_id": 666672917233594368, "created_time": "Tue Nov 17 17:43:04 +0000 2015", "geo": {"type": "Point", "coordinates": [52.63162952, -2.50153743]}}

35. {"original_text": "Something was missing... #PeopleOf pt14\n\n#Paris\n\nA simple portrait \nWaiting for the bus to arrive,\u2026 https://t.co/zsPQkq8HYC", "manual_label": "3", "hashtags": ["PeopleOf", "Paris"], "filtered_text": "someth missing pt14 a simpl portrait wait bu arrive", "original_id": 666814212157652992, "created_time": "Wed Nov 18 03:04:31 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

36. {"original_text": "He Can't multi-task! #waiting #agegap #HeOld!! @ Paris, France https://t.co/h4DQ6nYPES", "manual_label": "4", "hashtags": ["waiting", "agegap", "HeOld"], "filtered_text": "he can t multi task paris franc", "original_id": 666577332975587328, "created_time": "Tue Nov 17 11:23:15 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

37. {"original_text": "#Paris, we join you in #Cincinnati! #jesuisenterrasse #TousAuBistrot \ud83c\udf77 @ Columbia Township https://t.co/HemCzoPbU0", "manual_label": "3", "hashtags": ["Paris", "Cincinnati", "jesuisenterrasse",

"TousAuBistrot"], "filtered_text": "join columbia township 0", "original_id": 666781214338977792, "created_time": "Wed Nov 18 00:53:24 +0000 2015", "geo": {"type": "Point", "coordinates": [39.170236, -84.386322]}}

38. {"original_text": "At @ow2 #ow2con in Paris,to join a panel session about FOSS project governance", "manual_label": "1", "hashtags": ["ow2con"], "filtered_text": "at paris to join panel session foss project govern", "original_id": 666547936311033856, "created_time": "Tue Nov 17 09:26:26 +0000 2015", "geo": {"type": "Point", "coordinates": [48.84385337, 2.3880959]}}

39. {"original_text": "6. H\u00f8gmo\n7. ISIS\n8. Gratulerer\n9. Aftenposten\n10. Beirut\n\n2015/11/17 13:35 CET #trndnl https://t.co/6sjsp7X8c6", "manual_label": "6", "hashtags": ["trndnl"], "filtered_text": "6 h gmo 7 isi 8 gratuler 9 aftenposten 10 beirut 2015 11 17 13 35 cet", "original_id": 666596774098829312, "created_time": "Tue Nov 17 12:40:30 +0000 2015", "geo": {"type": "Point", "coordinates": [64.5565, 12.6654]}}

40. {"original_text": "The Guys @sheridanaveband \n#photoshoot #lvc #sheridanaveband #sheridanave #homies 4/6 @ Lebanon\u2026 https://t.co/yOh59iIeIV", "manual_label": "6", "hashtags": ["photoshoot", "lvc", "sheridanaveband", "sheridanave", "homies"], "filtered_text": "the guy 4 6 lebanon", "original_id": 666735918892486657, "created_time": "Tue Nov 17 21:53:25 +0000 2015", "geo": {"type": "Point", "coordinates": [40.33176802, -76.51479861]}}

41. {"original_text": "On Monday 16, 'Poutine' was Trending Topic in Paris for 6 hours: https://t.co/aGaBPMwTK8 #trndnl", "manual_label": "6", "hashtags": ["trndnl"], "filtered_text": "on monday 16 poutine trend topic pari 6 hours", "original_id": 666582618729771008, "created_time": "Tue Nov 17 11:44:15 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8569, 2.3412]}}

42. {"original_text": "#France 2 bomb Azam khan buffalo's?\nHey take away #AzamKhan leave \nprecious #Buffalo's\n#FranceAttacks #Hollande #Raqqa \n#SyrianWar #IraqWar", "manual_label": "2", "hashtags": ["France", "AzamKhan", "Buffalo", "FranceAttacks", "Hollande", "Raqqa", "SyrianWar", "IraqWar"], "filtered_text": "2 bomb azam khan buffalo s hey take away leav preciou s", "original_id": 666837687941246976, "created_time": "Wed Nov 18 04:37:48 +0000 2015", "geo": {"type": "Point", "coordinates": [15.8750453, 74.5025448]}}

43. {"original_text": "Want to work in #Lebanon, TN? View our latest opening: https://t.co/sMOOommg0n #Engineering #Veterans #Job #Jobs #Hiring", "manual_label": "5", "hashtags": ["Lebanon", "Engineering", "Veterans", "Job", "Jobs", "Hiring"], "filtered_text": "want work tn view latest opening", "original_id": 666686608729485312, "created_time": "Tue Nov 17 18:37:28 +0000 2015", "geo": {"type": "Point", "coordinates": [36.2081098, -86.2911024]}}

44. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/tpBKH3DHAz", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666807187386138624, "created_time": "Wed Nov 18 02:36:36 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

45. {"original_text": "Charlie Hebdo ..#charliehebdo #leman #uykusuz (at @Charlie_Hebdo_ in Paris, \u00cele-de-France) https://t.co/tVzypi9TCb https://t.co/25YGMNG8xh", "manual_label": "6", "hashtags": ["charliehebdo", "leman", "uykusuz"], "filtered_text": "charli hebdo at paris le de france https t co 25ygmng8xh", "original_id": 666604085634850816, "created_time": "Tue Nov 17 13:09:33 +0000 2015", "geo": {"type": "Point", "coordinates": [48.85888666, 2.37054651]}}

46. {"original_text": "Ajude quem precisa... S\u00f3 isso! \n\u2764\ufe0f\ud83d\ude4f\ud83c\udffc\n#mariana\n#ajudemarianamg\n#mg \n#paris\n#playforparis\n#Repost\u2026 https://t.co/vgSQgok5Tf", "manual_label": "3", "hashtags": ["mariana", "ajudemarianamg", "mg", "paris", "playforparis", "Repost"], "filtered_text": "ajud quem precisa s isso n m mg st tf", "original_id": 666486637090766848, "created_time": "Tue Nov 17 05:22:51 +0000 2015", "geo": {"type": "Point", "coordinates": [-2.6, -44.23333333]}}

47. {"original_text": "Les meilleurs #best #family #friends #parisian #paris #party #birthday debydebo026 deboamar55\u2026 https://t.co/hErqHnafPu", "manual_label": "4", "hashtags": ["best", "family", "friends", "parisian", "paris", "party", "birthday"], "filtered_text": "le meilleur debydebo026 deboamar55", "original_id": 666775862633611264, "created_time": "Wed Nov 18 00:32:08 +0000 2015", "geo": {"type": "Point", "coordinates": [48.87905849, 2.27838699]}}

48. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/hRJ4elUiC6", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666805227438211072, "created_time": "Wed Nov 18 02:28:49 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

49. {"original_text": "\u00c7a fait du bien de la voir aussi belle. #paris #tourEiffel #13thnovember\u2026 https://t.co/d3up5Tq78J", "manual_label": "2", "hashtags": ["paris", "tourEiffel", "13thnovember"], "filtered_text": "a fait du bien de la voir aussi belle", "original_id": 666716966078586880, "created_time": "Tue Nov 17 20:38:06 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86219263, 2.28795251]}}

50. {"original_text": "The affects of a tea bomb during your shake shift on Founders Day! #ApronOff #SSSO #Funky4\u2026 https://t.co/40KpBbW19f", "manual_label": "4",

"hashtags": ["ApronOff", "SSSO", "Funky4"], "filtered_text": "the affect tea bomb shake shift founder day", "original_id": 666765569354571776, "created_time": "Tue Nov 17 23:51:14 +0000 2015", "geo": {"type": "Point", "coordinates": [33.04068211, -96.7174629]}}

51. {"original_text": "Performance Food Group #Transportation : Driver (Route Delivery) Min $1100.00 a week!!!!! (#Lebanon, Tennessee) https://t.co/jhhooimkhI", "manual_label": "4", "hashtags": ["Transportation", "Lebanon"], "filtered_text": "perform food group driver rout delivery min 1100 00 week tennessee", "original_id": 666790503384375296, "created_time": "Wed Nov 18 01:30:19 +0000 2015", "geo": {"type": "Point", "coordinates": [35.6117453, -89.3575712]}}
52. {"original_text": "\ud83c\uddeb\ud83c\uddf7PARIS __ FRANCE\ud83d\udd35\u26aa\ufe0f\ud83d\udd34 #night #picture #river #toureiffel #photo #likes #igers #followme #bridge\u2026 https://t.co/ulWgt4Bxzp", "manual_label": "1", "hashtags": ["night", "picture", "river", "toureiffel", "photo", "likes", "igers", "followme", "bridge"], "filtered_text": "pari __ france f h g o id bxzp", "original_id": 666671105797697536, "created_time": "Tue Nov 17 17:35:52 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}
53. {"original_text": "Beautiful tribute outside the Pyramid tonight. #prayforparis #paris #csulb #vivelafrance @ California\u2026 https://t.co/aagdmDDp3n", "manual_label": "3", "hashtags": ["prayforparis", "paris", "csulb", "vivelafrance"], "filtered_text": "beauti tribut outsid pyramid tonight california", "original_id": 666549578452312064, "created_time": "Tue Nov 17 09:32:58 +0000 2015", "geo": {"type": "Point", "coordinates": [33.78138261, -118.1139261]}}
54. {"original_text": "@SenTomCotton @JakeTapper @CNN #ISIS #Paris #BreakingNews #ArrestBuhari explosion in Adamawa @PA @M_E_Adams @RT_com https://t.co/EYUSv5TYA9", "manual_label": "2", "hashtags": ["ISIS", "Paris", "BreakingNews", "ArrestBuhari"], "filtered_text": "explos adamawa https t co eyusv5tya9", "original_id": 666733851587289088, "created_time": "Tue Nov 17 21:45:12 +0000 2015", "geo": {"type": "Point", "coordinates": [6.43667, 7.48353]}}
55. {"original_text": "#paris @ Paris, France https://t.co/9x0SRvAmY5", "manual_label": "1", "hashtags": ["paris"], "filtered_text": "paris franc", "original_id": 666578005439983616, "created_time": "Tue Nov 17 11:25:55 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}
56. {"original_text": "#Paris\n#InGodWeTrust @ Paris, France https://t.co/yhxuVDFAS4", "manual_label": "3", "hashtags": ["Paris", "InGodWeTrust"], "filtered_text": "paris franc", "original_id": 666710094999089154, "created_time": "Tue Nov 17 20:10:48 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

57. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/UazWBYr7t3", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666807895657349120, "created_time": "Wed Nov 18 02:39:25 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

58. {"original_text": "Pop corn au homard\n#popcorn #streetfood #foodtruck #lobsterandco #lobster #homard @ Paris, France https://t.co/83o2ty5ZtH", "manual_label": "1", "hashtags": ["popcorn", "streetfood", "foodtruck", "lobsterandco", "lobster", "homard"], "filtered_text": "pop corn au homard paris franc", "original_id": 666707707089416192, "created_time": "Tue Nov 17 20:01:18 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

59. {"original_text": "Elle est magnifique \ud83c\uddeb\ud83c\uddf7\u2764\ufe0f #Paris #toureiffel #eiffeltower #Fluctuatnecmergitu @ Tour Eiffel https://t.co/Iscljj6kC3", "manual_label": "1", "hashtags": ["Paris", "toureiffel", "eiffeltower", "Fluctuatnecmergitu"], "filtered_text": "ell est magnifiqu e e tu tour eiff c3", "original_id": 666708623159140352, "created_time": "Tue Nov 17 20:04:57 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}

60. {"original_text": "\u201c@MazMHussain: one of the attackers, was 5 years old on 9/11.War on Terror & airstrikes no solution more creating multi-headed #Hydra #paris", "manual_label": "2", "hashtags": ["Hydra", "paris"], "filtered_text": "one attackers 5 year old 9 11 war terror amp airstrik solut creat multi head", "original_id": 666488201360580608, "created_time": "Tue Nov 17 05:29:04 +0000 2015", "geo": {"type": "Point", "coordinates": [52.37010055, 4.91085774]}}

61. {"original_text": "Can you recommend anyone for this #job? Co-op - Machining Engineering (Lebanon, MO) (Summer 2016 \u2013 May Start) - https://t.co/TDoOfaLo1P", "manual_label": "5", "hashtags": ["job"], "filtered_text": "can recommend anyon co op machin engin lebanon mo summer 2016 may start", "original_id": 666746337963044864, "created_time": "Tue Nov 17 22:34:49 +0000 2015", "geo": {"type": "Point", "coordinates": [37.6805967, -92.6637865]}}

62. {"original_text": "Come i #bambini vedono la #guerra e il #terrorismo #paris \n#nowar no terrorism we are all #brothers\u2026 https://t.co/lLQdwXjDe3", "manual_label": "2", "hashtags": ["bambini", "guerra", "terrorismo", "paris", "nowar", "brothers"], "filtered_text": "come vedono la e il terror", "original_id": 666531972546625536, "created_time": "Tue Nov 17 08:23:00 +0000 2015", "geo": {"type": "Point", "coordinates": [45.4667, 9.2]}}

63. {"original_text": "Heart attack #1DPL", "manual_label": "6", "hashtags": ["1DPL"], "filtered_text": "heart attack", "original_id": 666688900920950785, "created_time":

"Tue Nov 17 18:46:35 +0000 2015", "geo": {"type": "Point", "coordinates": [54.3841611, 19.8297019]}}

64. {"original_text": "Speechless... \ud83c\uddeb\ud83c\uddf7 #stadedefrance #Paris #day4 @ Stade de France \u2013 Saint-Denis https://t.co/oFLDoU1ZeO", "manual_label": "6", "hashtags": ["stadedefrance", "Paris", "day4"], "filtered_text": "speechless c y4 stade de franc saint deni eo", "original_id": 666526550867750912, "created_time": "Tue Nov 17 08:01:27 +0000 2015", "geo": {"type": "Point", "coordinates": [48.91777778, 2.35055556]}}
65. {"original_text": "#Stigmabase | AM - Copacabana LGBT parade honors Paris victims \u00a0 -\u00a0The Copacabana LGBT parade dampened by Paris\u2026 https://t.co/dpxyTeTDqV", "manual_label": "3", "hashtags": ["Stigmabase"], "filtered_text": "am copacabana lgbt parad honor pari victim the copacabana lgbt parad dampen paris", "original_id": 666750069777895425, "created_time": "Tue Nov 17 22:49:38 +0000 2015", "geo": {"type": "Point", "coordinates": [25.7231361, -80.20185547]}}
66. {"original_text": "Attack of the sea nettles. #montereybayaquarium #seanettle @ Monterey, CA, United States https://t.co/f37zf1YeRz", "manual_label": "1", "hashtags": ["montereybayaquarium", "seanettle"], "filtered_text": "attack sea nettles monterey ca unit state", "original_id": 666829583656886272, "created_time": "Wed Nov 18 04:05:36 +0000 2015", "geo": {"type": "Point", "coordinates": [36.6, -121.891]}}
67. {"original_text": "romyjanec : \"\ud83d\udc40\" https://t.co/5eUAMoIHa2 #beirut", "manual_label": "2", "hashtags": ["beirut"], "filtered_text": "romyjanec 2", "original_id": 666491100107022336, "created_time": "Tue Nov 17 05:40:35 +0000 2015", "geo": {"type": "Point", "coordinates": [33.89977205, 35.48200163]}}
68. {"original_text": "https://t.co/HMFo3KHMQo #beirut", "manual_label": "2", "hashtags": ["beirut"], "filtered_text": "", "original_id": 666576701699264512, "created_time": "Tue Nov 17 11:20:44 +0000 2015", "geo": {"type": "Point", "coordinates": [33.88682038, 35.5042974]}}
69. {"original_text": "6. #adtechNZ\n7. Simon Lusk\n8. ISIS\n9. Syria\n10. Paris\n\n2015/11/17 22:55 NZDT #trndnl https://t.co/WiiLMe8GDw", "manual_label": "6", "hashtags": ["adtechNZ", "trndnl"], "filtered_text": "6 7 simon lusk 8 isi 9 syria 10 pari 2015 11 17 22 55 nzdt", "original_id": 666556539281367040, "created_time": "Tue Nov 17 10:00:37 +0000 2015", "geo": {"type": "Point", "coordinates": [-43.5877, 170.3666]}}
70. {"original_text": "Milano non sar\u00e0 mai pi\u00f9 la stessa \ud83c\udf42 #prayforparis #prayfortheworld #milan #paris @ Piazza Fontana https://t.co/Rw5Re5jcti", "manual_label": "3", "hashtags": ["prayforparis", "prayfortheworld", "milan", "paris"], "filtered_text": "milano non sar mai pi la stessa d n piazza fontana",

"original_id": 666696222778880001, "created_time": "Tue Nov 17 19:15:40 +0000 2015", "geo": {"type": "Point", "coordinates": [45.46344098, 9.19406362]}}

71. {"original_text": "On Monday 16, 'Paris' was Trending Topic in Bristol for 9 hours: https://t.co/NVYZ4Fck2V #trndnl", "manual_label": "6", "hashtags": ["trndnl"], "filtered_text": "on monday 16 paris trend topic bristol 9 hours", "original_id": 666729791932862464, "created_time": "Tue Nov 17 21:29:04 +0000 2015", "geo": {"type": "Point", "coordinates": [51.4537, -2.5914]}}

72. {"original_text": "mhd_khatib : \"\u0628\u0627\u0644\u0631\u0648\u062d \u0644\u0627 \u062a\u0628\u0627\u0644\u064a \u0631\u0648\u062d \u0644\u0644\u063a\u0627\u0644\u064a\n \u0645\u0634\u064a \u0645\u0634\u064a \u0645\u0634\u064a \u0645\u0634\u064a\" https://t.co/ljYnt3jGqv #beirut", "manual_label": "6", "hashtags": ["beirut"], "filtered_text": "mhd_khatib", "original_id": 666531779231277056, "created_time": "Tue Nov 17 08:22:14 +0000 2015", "geo": {"type": "Point", "coordinates": [33.85139815, 35.52001102]}}

73. {"original_text": "What a great #Tribute l! Thank you so much #city of #LosAngeles for caring! #ThankYou #paris\u2026 https://t.co/xldcgUAZtc", "manual_label": "1", "hashtags": ["Tribute", "city", "LosAngeles", "ThankYou", "paris"], "filtered_text": "what great l thank much caring", "original_id": 666839400295305217, "created_time": "Wed Nov 18 04:44:36 +0000 2015", "geo": {"type": "Point", "coordinates": [34.0536, -118.243]}}

74. {"original_text": "Paris reste une f\u00eate ! #tousaubistrot #jesuisenterrasse #hommage #resistance @ La Marmite https://t.co/FhuZvDQH6B", "manual_label": "3", "hashtags": ["tousaubistrot", "jesuisenterrasse", "hommage", "resistance"], "filtered_text": "pari rest une f te la marmit", "original_id": 666750343192047616, "created_time": "Tue Nov 17 22:50:44 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8822899, 2.33939]}}

75. {"original_text": "Did you know that #E1matin was Trending Topic on Tuesday 17 for 4 hours in Paris? https://t.co/D67fdwWnro #trndnl", "manual_label": "6", "hashtags": ["E1matin", "trndnl"], "filtered_text": "did know trend topic tuesday 17 4 hour paris", "original_id": 666760091194949632, "created_time": "Tue Nov 17 23:29:28 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8569, 2.3412]}}

76. {"original_text": "Burj khalifa\ud83d\ude07\ud83d\ude07\ud83d\ude07 pray for paris #uae #dubai #france #paris #flag #lights #nightlife #nightlights @\u2026 https://t.co/VIseBzc4yF", "manual_label": "1", "hashtags": ["uae", "dubai", "france", "paris", "flag", "lights", "nightlife", "nightlights"], "filtered_text": "burj khalifa pray par u b n r l h ht 4yf", "original_id": 666680284654383105, "created_time": "Tue Nov 17 18:12:20 +0000 2015", "geo": {"type": "Point", "coordinates": [25.19588318, 55.27713686]}}

77. {"original_text": "Can you recommend anyone for this #job? 12K Plumber - https://t.co/L5kudeLRE4 #NettempsJobs #Paris, TN #Engineering #Hiring #CareerArc", "manual_label": "5", "hashtags": ["job", "NettempsJobs", "Paris", "Engineering", "Hiring", "CareerArc"], "filtered_text": "can recommend anyon 12k plumber tn", "original_id": 666686899868688385, "created_time": "Tue Nov 17 18:38:38 +0000 2015", "geo": {"type": "Point", "coordinates": [36.3020023, -88.3267107]}}

78. {"original_text": "Happy place \u2764\ud83c\uddeb\ud83c\uddf7\ufe0f\ud83d\ude4f\ud83c\udffd @fsparrows #Paris #coffeeshop @ Folks And Sparrows https://t.co/Rje4aDmUnO", "manual_label": "1", "hashtags": ["Paris", "coffeeshop"], "filtered_text": "happi place r shop folk and sparr muno", "original_id": 666602305756745728, "created_time": "Tue Nov 17 13:02:29 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8609505, 2.36869]}}

79. {"original_text": "Nice touch from saintraymondmusic last night. #paris @ Manchester, United Kingdom https://t.co/tbrSKMH1Qq", "manual_label": "1", "hashtags": ["paris"], "filtered_text": "nice touch saintraymondmus last night manchester unit kingdom", "original_id": 666595120284721153, "created_time": "Tue Nov 17 12:33:56 +0000 2015", "geo": {"type": "Point", "coordinates": [53.4667, -2.2333]}}

80. {"original_text": "#tuinglemoment #tuingletravel snapshot pf Paris, France. Our prayers are still with you. #paris\u2026 https://t.co/DZGIoYYUKV", "manual_label": "1", "hashtags": ["tuinglemoment", "tuingletravel", "paris"], "filtered_text": "snapshot pf paris france our prayer still you", "original_id": 666787098612129792, "created_time": "Wed Nov 18 01:16:47 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

81. {"original_text": "Seriously sounds like we're going to bomb #Syria \ud83d\udc4d \ud83d\udc4d \ud83d\udc4d \ud83d\udc4d #DestroyISIS", "manual_label": "2", "hashtags": ["Syria", "DestroyISIS"], "filtered_text": "serious sound like we r go bomb isi", "original_id": 666597523029680128, "created_time": "Tue Nov 17 12:43:28 +0000 2015", "geo": {"type": "Point", "coordinates": [52.38820222, -1.4841975]}}

82. {"original_text": "#pressday #optic2000 #lunettes #glasses #karllagerfeld #paris @ Rue de S\u00e9vign\u00e9 https://t.co/Q4s8MU9Jc3", "manual_label": "4", "hashtags": ["pressday", "optic2000", "lunettes", "glasses", "karllagerfeld", "paris"], "filtered_text": "rue de s vign", "original_id": 666592659172118528, "created_time": "Tue Nov 17 12:24:09 +0000 2015", "geo": {"type": "Point", "coordinates": [48.85680199, 2.3625774]}}

83. {"original_text": "Join the TeamHealth team! See our latest #job opening here: https://t.co/zu0YoLKjoH #LEBANON, KY #Hiring #CareerArc", "manual_label": "5", "hashtags": ["job", "LEBANON", "Hiring", "CareerArc"], "filtered_text": "join

teamhealth team see latest open here ky", "original_id": 666715330761256960, "created_time": "Tue Nov 17 20:31:36 +0000 2015", "geo": {"type": "Point", "coordinates": [37.5706264, -85.2606241]}}

84. {"original_text": "Paris / #paris \ud83d\udc99 @ Trocad\u00e9ro https://t.co/FDGx5SgQiE", "manual_label": "1", "hashtags": ["paris"], "filtered_text": "pari trocad ro e", "original_id": 666733911435821057, "created_time": "Tue Nov 17 21:45:26 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86219263, 2.28795251]}}

85. {"original_text": "Thurrock Mayor writes to Paris counterpart to express sympathies https://t.co/dbYCvY8i9p #Thurrock https://t.co/vllubi7FjG", "manual_label": "3", "hashtags": ["Thurrock"], "filtered_text": "thurrock mayor write pari counterpart express sympathi https t co vllubi7fjg", "original_id": 666676581830553600, "created_time": "Tue Nov 17 17:57:38 +0000 2015", "geo": {"type": "Point", "coordinates": [51.48344704, 0.36244053]}}

86. {"original_text": "\ud83d\uddfc\ud83d\ude22 #prayforparis #france #Paris #prayfortheworld #vsco @ Eiffel Tower https://t.co/iZI3blneCm", "manual_label": "3", "hashtags": ["prayforparis", "france", "Paris", "prayfortheworld", "vsco"], "filtered_text": "c l co eiffel tow cm", "original_id": 666806833273794560, "created_time": "Wed Nov 18 02:35:12 +0000 2015", "geo": {"type": "Point", "coordinates": [48.83505906, 2.38761335]}}

87. {"original_text": "#neoplasticism in Paris #centrepompidou #vsco #vscocam #afternoon @ Centre Pompidou https://t.co/j8pYrkYkmc", "manual_label": "1", "hashtags": ["neoplasticism", "centrepompidou", "vsco", "vscocam", "afternoon"], "filtered_text": "pari centr pompid", "original_id": 666809922504421376, "created_time": "Wed Nov 18 02:47:28 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8606796, 2.35198]}}

88. {"original_text": "'Paris' esteve na segunda-feira 16 como Assunto do Momento em Guarulhos durante 4 horas: https://t.co/N6TpnYdFDd #trndnl", "manual_label": "6", "hashtags": ["trndnl"], "filtered_text": "paris estev na segunda feira 16 como assunto momento em guarulho durant 4 horas", "original_id": 666763798531801090, "created_time": "Tue Nov 17 23:44:12 +0000 2015", "geo": {"type": "Point", "coordinates": [-23.444, -46.5078]}}

89. {"original_text": "\ud83c\udf19 #Paris \ud83d\udcf7 @gwwla @ Canal Saint-Martin https://t.co/FPicgi1JpL", "manual_label": "1", "hashtags": ["Paris"], "filtered_text": "la canal saint marti pl", "original_id": 666528224894181376, "created_time": "Tue Nov 17 08:08:06 +0000 2015", "geo": {"type": "Point", "coordinates": [48.874892, 2.363386]}}

90. {"original_text": "#webstagram #europe #emirates #rider #young #uae #usa #instadubai #paris #jumeirah #jumeirahbeach\u2026 https://t.co/UKBhwX3FDs",

"manual_label": "1", "hashtags": ["webstagram", "europe", "emirates", "rider", "young", "uae", "usa", "instadubai", "paris", "jumeirah", "jumeirahbeach"], "filtered_text": "", "original_id": 666522418282819584, "created_time": "Tue Nov 17 07:45:02 +0000 2015", "geo": {"type": "Point", "coordinates": [24.36888174, 54.52090481]}}

91. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/5VLxlgkvPf", "manual_label": "6", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666808131565916161, "created_time": "Wed Nov 18 02:40:21 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

92. {"original_text": "#morning #session #yoga #peace #paris @ Palestra Montecatini https://t.co/vAEZtf0GGK", "manual_label": "4", "hashtags": ["morning", "session", "yoga", "peace", "paris"], "filtered_text": "palestra montecatini", "original_id": 666544402060345345, "created_time": "Tue Nov 17 09:12:23 +0000 2015", "geo": {"type": "Point", "coordinates": [43.878231, 10.7789202]}}

93. {"original_text": "#Repost paris.38 with repostapp.\n\u30fb\u30fb\u30fb\nMaquillaje express #tachi jackelyn.gaibor \ud83d\ude1a @ Tachi https://t.co/16xt6d4tti", "manual_label": "4", "hashtags": ["Repost", "tachi"], "filtered_text": "paris 38 repostapp maquillaj express jackelyn gaibor tachi", "original_id": 666723793369346048, "created_time": "Tue Nov 17 21:05:14 +0000 2015", "geo": {"type": "Point", "coordinates": [40.422451, -3.66694]}}

94. {"original_text": "Les petits #parisiens dessinent #Paris #Libre @ Le Bataclan https://t.co/zuRGtp1OhV", "manual_label": "3", "hashtags": ["parisiens", "Paris", "Libre"], "filtered_text": "le petit dessin le bataclan", "original_id": 666602201695956992, "created_time": "Tue Nov 17 13:02:04 +0000 2015", "geo": {"type": "Point", "coordinates": [48.863121, 2.3708701]}}

95. {"original_text": "#Paris zabouartist #streetart #arturbain #urbanart #paint #artderue #zabou #zabouartist @ Montmartre https://t.co/p6qFHp7n5i", "manual_label": "1", "hashtags": ["Paris", "streetart", "arturbain", "urbanart", "paint", "artderue", "zabou", "zabouartist"], "filtered_text": "zabouartist montmartr", "original_id": 666731331443789824, "created_time": "Tue Nov 17 21:35:11 +0000 2015", "geo": {"type": "Point", "coordinates": [48.88694444, 2.34111111]}}

96. {"original_text": "#TousAuBistrot \n\ud83c\uddeb\ud83c\uddf7 @ Paris 10\u00e8me https://t.co/c1LFNpWwLF", "manual_label": "1", "hashtags": ["TousAuBistrot"], "filtered_text": "pari 10 m lf", "original_id": 666718740713598976, "created_time": "Tue Nov 17 20:45:09 +0000 2015", "geo": {"type": "Point", "coordinates": [48.87666201, 2.36878429]}}

97. {"original_text": "#Merci #TankYou \ud83c\uddeb\ud83c\uddf7\ud83c\uddec\ud83c\udde7\u2764\ufe0f @ Paris, France https://t.co/gfO8FdD6l4", "manual_label": "3", "hashtags": ["Merci", "TankYou"], "filtered_text": "paris fra d6l4", "original_id": 666708167502393346, "created_time": "Tue Nov 17 20:03:08 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

98. {"original_text": "#hello #adele #adelehello #parisstreet #frenchlife\ud83c\uddeb\ud83c\uddf7 #paris #metro @ Paris, France https://t.co/FJ9bHQuF58", "manual_label": "1", "hashtags": ["hello", "adele", "adelehello", "parisstreet", "frenchlife", "paris", "metro"], "filtered_text": "ro paris franc 58", "original_id": 666772840155344896, "created_time": "Wed Nov 18 00:20:07 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

99. {"original_text": "Beautiful fall days... In studio \ud83d\ude02\ud83d\uddfb\ud83c\udf42\ud83c\udfac\ud83c\udfa5 New film coming soon! #setlife #reddragon #paris #studio\u2026 https://t.co/tNJtl5JwMf", "manual_label": "1", "hashtags": ["setlife", "reddragon", "paris", "studio"], "filtered_text": "beauti fall days in studio new film come r p tu 5jwmf", "original_id": 666689193293123584, "created_time": "Tue Nov 17 18:47:44 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8904, 2.3426]}}

100. {"original_text": "6. #MTVStars\n7. Simon Lusk\n8. ISIS\n9. Paris\n10. Ross Taylor\n\n2015/11/18 01:15 NZDT #trndnl https://t.co/WiiLMe8GDw", "manual_label": "6", "hashtags": ["MTVStars", "trndnl"], "filtered_text": "6 7 simon lusk 8 isi 9 pari 10 ross taylor 2015 11 18 01 15 nzdt", "original_id": 666591781677457409, "created_time": "Tue Nov 17 12:20:40 +0000 2015", "geo": {"type": "Point", "coordinates": [-43.5877, 170.3666]}}

101. {"original_text": "'Paris' appeared on Sunday 16 at the 2nd place in the Top20 of Australia's Trends: https://t.co/w1EZoZzuB4 #trndnl", "manual_label": "6", "hashtags": ["trndnl"], "filtered_text": "paris appear sunday 16 2nd place top20 australia trends", "original_id": 666537076137918464, "created_time": "Tue Nov 17 08:43:17 +0000 2015", "geo": {"type": "Point", "coordinates": [-24.9162, 133.3931]}}

102. {"original_text": "Nights R a blessing \ud83d\ude4f\ud83c\udffdwhen I'm spending them like this... Right @ my feet\u2764\ufe0f #Paris & #knicks \ud83d\udc08\u2026 https://t.co/8MBuNKVDnZ", "manual_label": "1", "hashtags": ["Paris", "knicks"], "filtered_text": "night r bless when i m spend like this right feet amp ks dnz", "original_id": 666791692381642753, "created_time": "Wed Nov 18 01:35:02 +0000 2015", "geo": {"type": "Point", "coordinates": [40.7142, -74.0064]}}

103. {"original_text": "#paris #libert\u00e9 #toujours #gauloises @ Amman, Jordan https://t.co/hZCIZMzoTF", "manual_label": "1", "hashtags": ["paris", "libert\u00e9",

"toujours", "gauloises"], "filtered_text": "amman jordan", "original_id": 666713675202342912, "created_time": "Tue Nov 17 20:25:01 +0000 2015", "geo": {"type": "Point", "coordinates": [31.95, 35.9333]}}

104. {"original_text": "LE RUN DU MARDI | HOMMAGE \u00c0 PARIS \ud83c\uddeb\ud83c\uddf7\u2728\n#rundumardi #runners #running #crossfit #streetworkout #paris\u2026 https://t.co/D5VagAjWKj", "manual_label": "1", "hashtags": ["rundumardi", "runners", "running", "crossfit", "streetworkout", "paris"], "filtered_text": "le run du mardi hommag pari d r n u kj", "original_id": 666746523078774784, "created_time": "Tue Nov 17 22:35:33 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8831978, 2.3697503]}}

105. {"original_text": "\"@FemiDlive: Retweeted Mr Tomide (@Tomyboiz): BEEAKING: BOMB BLAST KILLS OVER 30 IN YOLA,ADAMAWA STATE #PrayForNigeria\" @RadioPaparazi", "manual_label": "2", "hashtags": ["PrayForNigeria"], "filtered_text": "retweet mr tomid beeaking bomb blast kill over 30 in yola adamawa state", "original_id": 666710478962499584, "created_time": "Tue Nov 17 20:12:19 +0000 2015", "geo": {"type": "Point", "coordinates": [6.67097, 3.32569]}}

106. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/EJTGFb3ET5", "manual_label": "6", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666810464211333120, "created_time": "Wed Nov 18 02:49:38 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

107. {"original_text": "Want to work in #Lebanon, Tennessee? View our latest opening: https://t.co/YON7lcRloq #SupplyChain #Job #Jobs #Hiring", "manual_label": "5", "hashtags": ["Lebanon", "SupplyChain", "Job", "Jobs", "Hiring"], "filtered_text": "want work tennessee view latest opening", "original_id": 666721403312345089, "created_time": "Tue Nov 17 20:55:44 +0000 2015", "geo": {"type": "Point", "coordinates": [35.6117453, -89.3575712]}}

108. {"original_text": "#garebibliothequefrancoismiterand #paris12 #rerc #metro #ligne14 #empty #iphone6plus #paris @ Gare\u2026 https://t.co/psEVkpAgix", "manual_label": "1", "hashtags": ["garebibliothequefrancoismiterand", "paris12", "rerc", "metro", "ligne14", "empty", "iphone6plus", "paris"], "filtered_text": "gare", "original_id": 666713899396300801, "created_time": "Tue Nov 17 20:25:55 +0000 2015", "geo": {"type": "Point", "coordinates": [48.87798001, 2.32879169]}}

109. {"original_text": "Am in love with you \ud83c\udf39\u2764\ufe0f\n#leurope #france #paris #champselys\u00e9es #laduree @\u2026 https://t.co/ZZbdDxMH7O", "manual_label": "4", "hashtags": ["leurope", "france", "paris", "champselys\u00e9es", "laduree"], "filtered_text": "am love e e e o", "original_id": 666552508056100864,

"created_time": "Tue Nov 17 09:44:36 +0000 2015", "geo": {"type": "Point", "coordinates": [48.87102288, 2.30328988]}}

110. {"original_text": "Merci #france #freedom #candles #mourning #prayforparis #paris @ Place de la R\u00e9publique https://t.co/MC8wteYg8h", "manual_label": "3", "hashtags": ["france", "freedom", "candles", "mourning", "prayforparis", "paris"], "filtered_text": "merci place de la r publiqu", "original_id": 666785249569976320, "created_time": "Wed Nov 18 01:09:26 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86727778, 2.36405556]}}
111. {"original_text": "Where is Paris on the map? Play the game at https://t.co/t8uzWiaOqr #Paris", "manual_label": "6", "hashtags": ["Paris"], "filtered_text": "where pari map play game", "original_id": 666672160409296896, "created_time": "Tue Nov 17 17:40:03 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.35099]}}
112. {"original_text": "Can you recommend anyone for this #job? ASST STORE MGR, 3305 PARIS RD, CHALMETTE LA - https://t.co/vagLRHYsih #Chalmette, LA #Veterans", "manual_label": "5", "hashtags": ["job", "Chalmette", "Veterans"], "filtered_text": "can recommend anyon asst store mgr 3305 pari rd chalmett la la", "original_id": 666710845229920256, "created_time": "Tue Nov 17 20:13:47 +0000 2015", "geo": {"type": "Point", "coordinates": [29.9492817, -89.9593078]}}
113. {"original_text": "Good morning #Paris from #Sonoma, #California #Wine Country.https://t.co/Rhxa0Q1vx4", "manual_label": "1", "hashtags": ["Paris", "Sonoma", "California", "Wine"], "filtered_text": "good morn country https t co rhxa0q1vx4", "original_id": 666485514363514885, "created_time": "Tue Nov 17 05:18:23 +0000 2015", "geo": {"type": "Point", "coordinates": [38.28308279, -122.46323279]}}
114. {"original_text": "L'amour vaincra.\n#Paris #RueDeCharonne #11th #Attack #Message #Wall #Stranger #Post #Respect #Peace\u2026 https://t.co/GRoTbOPncS", "manual_label": "3", "hashtags": ["Paris", "RueDeCharonne", "11th", "Attack", "Message", "Wall", "Stranger", "Post", "Respect", "Peace"], "filtered_text": "l amour vaincra", "original_id": 666748981104541698, "created_time": "Tue Nov 17 22:45:19 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}
115. {"original_text": "Burj Khalifa #dubai #uae #mydubai #dxb #lovedubai #burjkhalifa #prayersforparis #paris @ Burj\u2026 https://t.co/ed0xAGfC4X", "manual_label": "3", "hashtags": ["dubai", "uae", "mydubai", "dxb", "lovedubai", "burjkhalifa", "prayersforparis", "paris"], "filtered_text": "burj khalifa burj", "original_id": 666690699623014404, "created_time": "Tue Nov 17 18:53:43 +0000 2015", "geo": {"type": "Point", "coordinates": [25.19322829, 55.27365642]}}
116. {"original_text": "Barcelona tribute to Paris #tribute #memory #paris #france #barcelona #spain #espana #peace #respect\u2026 https://t.co/VYT2tTbF11",

"manual_label": "3", "hashtags": ["tribute", "memory", "paris", "france", "barcelona", "spain", "espana", "peace", "respect"], "filtered_text": "barcelona tribut pari", "original_id": 666532727382913024, "created_time": "Tue Nov 17 08:26:00 +0000 2015", "geo": {"type": "Point", "coordinates": [41.38981595, 2.16481917]}}

117. {"original_text": "The shortest Trends on Monday 16 in Paris had only 4 characters long: https://t.co/aGaBPMOuBG #trndnl", "manual_label": "6", "hashtags": ["trndnl"], "filtered_text": "the shortest trend monday 16 pari 4 charact long", "original_id": 666699673466335232, "created_time": "Tue Nov 17 19:29:23 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8569, 2.3412]}}

118. {"original_text": "#artselfie #paris #frenchlife\ud83c\uddeb\ud83c\uddf7 #frenchie #frenchart @ Mus\u00e9e d'Orsay (officiel) https://t.co/VHdilubhht", "manual_label": "1", "hashtags": ["artselfie", "paris", "frenchlife", "frenchie", "frenchart"], "filtered_text": "rt mus d orsay officiel ht", "original_id": 666801013005905920, "created_time": "Wed Nov 18 02:12:04 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86025591, 2.32602851]}}

119. {"original_text": "leoniefreiji : \"Autumn leaves. #favorite #nightOut #london #LetsGo #Soon \ud83d\ude0c\ud83d\ude06\" https://t.co/0r2rwDSGXi #beirut", "manual_label": "1", "hashtags": ["favorite", "nightOut", "london", "LetsGo", "Soon", "beirut"], "filtered_text": "leoniefreiji autumn leaves x ut", "original_id": 666677411661127680, "created_time": "Tue Nov 17 18:00:55 +0000 2015", "geo": {"type": "Point", "coordinates": [33.8948614, 35.51283186]}}

120. {"original_text": "Exclusive Sketches: Designers Stand Behind Paris Amidst Tragedy https://t.co/6TGW661lV1 #Fashion https://t.co/VSzoke8OfN", "manual_label": "2", "hashtags": ["Fashion"], "filtered_text": "exclus sketches design stand behind pari amidst tragedi https t co vszoke8ofn", "original_id": 666681024760295424, "created_time": "Tue Nov 17 18:15:17 +0000 2015", "geo": {"type": "Point", "coordinates": [54.9026209, -2.48664781]}}

121. {"original_text": "shopping addiction ! yikes #boots #cowgirl #country #vegas @ Paris Las Vegas Hotel & Casino https://t.co/XcLkOukOmJ", "manual_label": "1", "hashtags": ["boots", "cowgirl", "country", "vegas"], "filtered_text": "shop addict yike pari la vega hotel amp casino", "original_id": 666739385824595968, "created_time": "Tue Nov 17 22:07:11 +0000 2015", "geo": {"type": "Point", "coordinates": [36.11235778, -115.17147064]}}

122. {"original_text": "#MoulinRouge #\u00c7aCestParis #Paris #ParisByNight #TonyGomez #ParisLaNuit #jaimeparis #LeMoulinRouge\u2026 https://t.co/jkHUc4CEUf", "manual_label": "1", "hashtags": ["MoulinRouge", "\u00c7aCestParis", "Paris", "ParisByNight", "TonyGomez", "ParisLaNuit", "jaimeparis", "LeMoulinRouge"], "filtered_text": "", "original_id":

666762121833816067, "created_time": "Tue Nov 17 23:37:32 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8840218, 2.3326199]}}

123. {"original_text": "CGI #internship #Job: Entry Level Business Analyst - Campus Recruiting (#Lebanon, VA) https://t.co/cuACD6ed5t #Jobs #Hiring #CareerArc", "manual_label": "5", "hashtags": ["internship", "Job", "Lebanon", "Jobs", "Hiring", "CareerArc"], "filtered_text": "cgi entri level busi analyst campu recruit va", "original_id": 666603243250692097, "created_time": "Tue Nov 17 13:06:12 +0000 2015", "geo": {"type": "Point", "coordinates": [36.900942, -82.0801309]}}

124. {"original_text": "The Parisian way of life can't be destroyed! #paris #parisparis #pariscity #city #citylife #parislife\u2026 https://t.co/JG3SzBGVVo", "manual_label": "4", "hashtags": ["paris", "parisparis", "pariscity", "city", "citylife", "parislife"], "filtered_text": "the parisian way life can t destroyed", "original_id": 666578550074380288, "created_time": "Tue Nov 17 11:28:05 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

125. {"original_text": "Me and Ele my Princess at #prpawards @ Cafe De Paris -London https://t.co/EZLu3nauqy", "manual_label": "1", "hashtags": ["prpawards"], "filtered_text": "me ele princess cafe de pari london", "original_id": 666778167206481920, "created_time": "Wed Nov 18 00:41:17 +0000 2015", "geo": {"type": "Point", "coordinates": [51.6765441, -0.0331088]}}

126. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/koyLlxhukM", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666806858221400064, "created_time": "Wed Nov 18 02:35:18 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

127. {"original_text": "French flag at half mast #france #jesuisparis #paris #melbourne visitmelbourne @cityofmelbourne @\u2026 https://t.co/171gzndfPz", "manual_label": "1", "hashtags": ["france", "jesuisparis", "paris", "melbourne"], "filtered_text": "french flag half mast visitmelbourn", "original_id": 666509685588164608, "created_time": "Tue Nov 17 06:54:26 +0000 2015", "geo": {"type": "Point", "coordinates": [-37.81789122, 144.96836333]}}

128. {"original_text": "L'Or\u00e9al Paris India, @lorealparisin is now trending in #Mumbai https://t.co/SGugy08vLO", "manual_label": "6", "hashtags": ["Mumbai"], "filtered_text": "l or al pari india trend", "original_id": 666544219129970688, "created_time": "Tue Nov 17 09:11:40 +0000 2015", "geo": {"type": "Point", "coordinates": [19.0728, 72.8826]}}

129. {"original_text": "[Stage] Stage d\u00e9veloppement Web H/F chez Codify - Paris\u00a0(Paris) : https://t.co/skf1fKPWjg #stage #informatique", "manual_label": "6", "hashtags": ["stage", "informatique"], "filtered_text": "stage stage d velopp web

h f chez codifi pari paris", "original_id": 666589122274947073, "created_time": "Tue Nov 17 12:10:06 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8772058, 2.3520399]}}

130. {"original_text": "How your right to know about government decisions and public money is under attack https://t.co/0FT03L1qWY #Rutland https://t.co/tNT0qPFA8c", "manual_label": "6", "hashtags": ["Rutland"], "filtered_text": "how right know govern decis public money attack https t co tnt0qpfa8c", "original_id": 666673498467667968, "created_time": "Tue Nov 17 17:45:22 +0000 2015", "geo": {"type": "Point", "coordinates": [52.65048587, -0.63751549]}}

131. {"original_text": "6. Fran\u00e7a\n7. #PrayForSyria\n8. #2DaysTilIKWYDLS\n9. Jessica\n10. Paris\n\n2015/11/17 07:55 WET #trndnl https://t.co/uLzQlByvJf", "manual_label": "6", "hashtags": ["PrayForSyria", "2DaysTilIKWYDLS", "trndnl"], "filtered_text": "6 fran a 7 8 9 jessica 10 pari 2015 11 17 07 55 wet", "original_id": 666526330469511168, "created_time": "Tue Nov 17 08:00:35 +0000 2015", "geo": {"type": "Point", "coordinates": [38.9901, -9.1413]}}

132. {"original_text": "#mma Holm executed flawless plan to crush Rousey - Lebanon Democrat: Food World News Holm\u2026 https://t.co/UQQFDsEDUE https://t.co/QYO5zGQkPj", "manual_label": "6", "hashtags": ["mma"], "filtered_text": "holm execut flawless plan crush rousey lebanon democrat food world news holm", "original_id": 666723626616406017, "created_time": "Tue Nov 17 21:04:34 +0000 2015", "geo": {"type": "Point", "coordinates": [36.10774555, -115.15019404]}}

133. {"original_text": "6. #MTVStars\n7. Simon Lusk\n8. ISIS\n9. Syria\n10. Paris\n\n2015/11/17 23:15 NZDT #trndnl https://t.co/WiiLMe8GDw", "manual_label": "6", "hashtags": ["MTVStars", "trndnl"], "filtered_text": "6 7 simon lusk 8 isi 9 syria 10 pari 2015 11 17 23 15 nzdt", "original_id": 666561582286024704, "created_time": "Tue Nov 17 10:20:39 +0000 2015", "geo": {"type": "Point", "coordinates": [-43.5877, 170.3666]}}

134. {"original_text": "Conf\u00e9rence Efashion 2015 #CCMefashion #conference #paris #efashion #fashion #stanleystella\u2026 https://t.co/VY1NRNPZyq", "manual_label": "1", "hashtags": ["CCMefashion", "conference", "paris", "efashion", "fashion", "stanleystella"], "filtered_text": "conf renc efashion 2015", "original_id": 666547700209287168, "created_time": "Tue Nov 17 09:25:30 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8699291, 2.33093791]}}

135. {"original_text": "#DirectLive : #TourEiffel #EiffelTower #Paris \ud83c\udf1e #ILoveParis #WeLoveParis #France @ Tour Eiffel https://t.co/0X8MHjk0VA", "manual_label": "1", "hashtags": ["DirectLive", "TourEiffel", "EiffelTower", "Paris", "ILoveParis", "WeLoveParis", "France"], "filtered_text": "e tour eiffel a", "original_id": 666674197092048897, "created_time": "Tue Nov 17 17:48:09 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}

136. {"original_text": "My 10 seconds of fame with Karl Stefanovic in #Paris @thetodayshow @ Place de la Republique https://t.co/1k4vpicvYZ", "manual_label": "1", "hashtags": ["Paris"], "filtered_text": "my 10 second fame karl stefanov place de la republiqu", "original_id": 666549888667271168, "created_time": "Tue Nov 17 09:34:12 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86746503, 2.36418438]}}

137. {"original_text": "Candles were lit today at a Vigil for #NohemiGonzales (victim of the terrorist attacks in Paris) at\u2026 https://t.co/yeJ410TDmj", "manual_label": "3", "hashtags": ["NohemiGonzales"], "filtered_text": "candl lit today vigil victim terrorist attack paris at", "original_id": 666843347554693120, "created_time": "Wed Nov 18 05:00:18 +0000 2015", "geo": {"type": "Point", "coordinates": [33.9792442, -118.0440216]}}

138. {"original_text": "@mouv dans le m\u00e9tro! #HipHopNeverStop @ Paris, France https://t.co/hgfuI7Yn9Q", "manual_label": "1", "hashtags": ["HipHopNeverStop"], "filtered_text": "dan le m tro paris franc", "original_id": 666598096734941184, "created_time": "Tue Nov 17 12:45:45 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

139. {"original_text": "C'est tous!! They have weapons. Fuck them, We have champagne! #francais #paris #charliehebdo @ New\u2026 https://t.co/2G5mVhDvBA", "manual_label": "6", "hashtags": ["francais", "paris", "charliehebdo"], "filtered_text": "c est tous they weapons fuck them we champagne new", "original_id": 666831417104551936, "created_time": "Wed Nov 18 04:12:53 +0000 2015", "geo": {"type": "Point", "coordinates": [40.7142, -74.0064]}}

140. {"original_text": "Man jailed for life for savage attack https://t.co/iHM2MrsUdi #Cumbria https://t.co/W0VVSgKFCy", "manual_label": "2", "hashtags": ["Cumbria"], "filtered_text": "man jail life savag attack https t co w0vvsgkfci", "original_id": 666604801405227012, "created_time": "Tue Nov 17 13:12:24 +0000 2015", "geo": {"type": "Point", "coordinates": [54.89453982, -2.93380223]}}

141. {"original_text": "#art #france #frenchart #streetlife #streetart @ Paris, France https://t.co/S5IMz78Uyc", "manual_label": "1", "hashtags": ["art", "france", "frenchart", "streetlife", "streetart"], "filtered_text": "paris franc", "original_id": 666811432399425536, "created_time": "Wed Nov 18 02:53:28 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

142. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/Umte2O5b7N", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666810873831292929, "created_time": "Wed Nov 18 02:51:15 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

143. {"original_text": "#dessert #dessertporn #parisjetaime #parisstreet #frenchlife\ud83c\uddeb\ud83c\uddf7 #frenchie #mango #mangue @ Paris, France https://t.co/DKdJoL2xKr", "manual_label": "1", "hashtags": ["dessert", "dessertporn", "parisjetaime", "parisstreet", "frenchlife", "frenchie", "mango", "mangue"], "filtered_text": "g ue paris franc kr", "original_id": 666769331930439681, "created_time": "Wed Nov 18 00:06:11 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

144. {"original_text": "I needed it I guess. #Bataclan #prayforparis #mourning @ Le Bataclan https://t.co/F8ZnZTlMHC", "manual_label": "3", "hashtags": ["Bataclan", "prayforparis", "mourning"], "filtered_text": "i need i guess le bataclan", "original_id": 666732142672666625, "created_time": "Tue Nov 17 21:38:24 +0000 2015", "geo": {"type": "Point", "coordinates": [48.863121, 2.3708701]}}

145. {"original_text": ".\n.\n\ud83d\udc6b People and colors are coming back in #Paris !\n.\n\u26ea\ufe0f Sur le Parvis de Notre-Dame !\n.\n\ud83d\udcda Allez...le\u2026 https://t.co/0wLKsIqLuH", "manual_label": "1", "hashtags": ["Paris"], "filtered_text": "peopl color come back sur le parvi de notre dam allez l uh", "original_id": 666685135803146240, "created_time": "Tue Nov 17 18:31:37 +0000 2015", "geo": {"type": "Point", "coordinates": [48.85361, 2.34806]}}

146. {"original_text": "WORD\n#dontknowwhatyouvegottilitsgone @ Lebanon, Maine https://t.co/gYFc1ongqG", "manual_label": "6", "hashtags": ["dontknowwhatyouvegottilitsgone"], "filtered_text": "word lebanon main", "original_id": 666788505545129984, "created_time": "Wed Nov 18 01:22:22 +0000 2015", "geo": {"type": "Point", "coordinates": [43.3944, -70.8514]}}

147. {"original_text": "C\u00e9r\u00e9monie de remise des dipl\u00f4mes #MBA #SMS (@ Stade Jean Bouin in Paris, \u00cele-de-France) https://t.co/EOShPVLsy0 https://t.co/5LwLpLS1bZ", "manual_label": "6", "hashtags": ["MBA", "SMS"], "filtered_text": "c r moni de remis de dipl m stade jean bouin paris le de france https t co 5lwlpls1bz", "original_id": 666673921702420480, "created_time": "Tue Nov 17 17:47:03 +0000 2015", "geo": {"type": "Point", "coordinates": [48.84318, 2.252919]}}

148. {"original_text": "#Transportation #Job in #PARIS, MO: Driver Helper at UPS https://t.co/OO9cdMn8FJ #Jobs #Hiring #CareerArc", "manual_label": "5", "hashtags": ["Transportation", "Job", "PARIS", "Jobs", "Hiring", "CareerArc"], "filtered_text": "mo driver helper up", "original_id": 666671439219560449, "created_time": "Tue Nov 17 17:37:11 +0000 2015", "geo": {"type": "Point", "coordinates": [39.4808721, -92.0012811]}}

149. {"original_text": "Bonne nuit! On change pour le meilleur chaque jour @leandrojusten vipmodelsparis mint_mgmt_nyc #paris\u2026 https://t.co/2ChYU6g4cD", "manual_label": "6", "hashtags": ["paris"], "filtered_text": "bonn nuit

on chang pour le meilleur chaqu jour vipmodelspari mint_mgmt_nyc", "original_id": 666747768036167680, "created_time": "Tue Nov 17 22:40:30 +0000 2015", "geo": {"type": "Point", "coordinates": [40.7142, -74.0064]}}

150. {"original_text": "Shy'm - Bercy!!! THA BOMB!!! #shym #bercy #paradoxaletour @ AccorHotels Arena https://t.co/igBVkNLiDk", "manual_label": "1", "hashtags": ["shym", "bercy", "paradoxaletour"], "filtered_text": "shy m bercy tha bomb accorhotel arena", "original_id": 666758481664315392, "created_time": "Tue Nov 17 23:23:04 +0000 2015", "geo": {"type": "Point", "coordinates": [48.83875005, 2.37919344]}}

151. {"original_text": "The museums opened today. Checking out the #museedorsay ! #Monet #koeurotrip #Paris @ Mus\u00e9e d'Orsay\u2026 https://t.co/HCParJc6a6", "manual_label": "1", "hashtags": ["museedorsay", "Monet", "koeurotrip", "Paris"], "filtered_text": "the museum open today check mus d orsay", "original_id": 666704051753590784, "created_time": "Tue Nov 17 19:46:47 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86025591, 2.32602851]}}

152. {"original_text": "#tb photo. . Persian #scouts #boysscouts @ Paris, \u00cele-de-France, France https://t.co/ymRWG3fsLD", "manual_label": "1", "hashtags": ["tb", "scouts", "boysscouts"], "filtered_text": "photo persian paris le de france franc", "original_id": 666818394767892480, "created_time": "Wed Nov 18 03:21:08 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

153. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/GQNWSMr4GQ", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666802872248479744, "created_time": "Wed Nov 18 02:19:27 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

154. {"original_text": "#frenchart #frenchlife\ud83c\uddeb\ud83c\uddf7 #paris #museum @ Mus\u00e9e d'Orsay (officiel) https://t.co/TKzZsaBPtS", "manual_label": "1", "hashtags": ["frenchart", "frenchlife", "paris", "museum"], "filtered_text": "um mus d orsay officiel ts", "original_id": 666801257139609600, "created_time": "Wed Nov 18 02:13:02 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86025591, 2.32602851]}}

155. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/hDUQMzMeno", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666805586722336768, "created_time": "Wed Nov 18 02:30:15 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

156. {"original_text": "JE SUIS PARIS \n\n#jesuisparis #paris #francia #france #instdaily #eiffel #eiffeltower #tower\u2026 https://t.co/c6vlpH3qL4", "manual_label": "1", "hashtags": ["jesuisparis", "paris", "francia", "france", "instdaily", "eiffel", "eiffeltower", "tower"], "filtered_text": "je sui pari", "original_id": 666525575222194176, "created_time": "Tue Nov 17 07:57:35 +0000 2015", "geo": {"type": "Point", "coordinates": [48.9012525, 2.3346539]}}

157. {"original_text": "#prayforparis #jesuisparis #prayforeveryone @ Paris - Tour Eiffel https://t.co/elYW0EVQ1M", "manual_label": "3", "hashtags": ["prayforparis", "jesuisparis", "prayforeveryone"], "filtered_text": "pari tour eiffel", "original_id": 666590151024381952, "created_time": "Tue Nov 17 12:14:11 +0000 2015", "geo": {"type": "Point", "coordinates": [43.93735683, 4.81249409]}}

158. {"original_text": "Nous on sort ! #parisattacks #resiste @ Parc Des Expositions (Paris Expo) https://t.co/6rTO9PKrBM", "manual_label": "1", "hashtags": ["parisattacks", "resiste"], "filtered_text": "nou sort parc de exposit pari expo", "original_id": 666691197839241219, "created_time": "Tue Nov 17 18:55:42 +0000 2015", "geo": {"type": "Point", "coordinates": [48.83067065, 2.28800222]}}

159. {"original_text": "#flag #halfmast #paris #beruit #inmemorium #eisenhowermedicalcenter #ranchomirage #palmsprings\u2026 https://t.co/IpdgHvCENg", "manual_label": "2", "hashtags": ["flag", "halfmast", "paris", "beruit", "inmemorium", "eisenhowermedicalcenter", "ranchomirage", "palmsprings"], "filtered_text": "", "original_id": 666516330330374144, "created_time": "Tue Nov 17 07:20:51 +0000 2015", "geo": {"type": "Point", "coordinates": [33.76337319, -116.4055836]}}

160. {"original_text": "Cheese Bomb Burger atm. Yumyum! #cheeseoverload #metime @ HID Burgers Mandaluyong https://t.co/eFoNvnJ5EL", "manual_label": "1", "hashtags": ["cheeseoverload", "metime"], "filtered_text": "chees bomb burger atm yumyum hid burger mandaluyong", "original_id": 666538894465208320, "created_time": "Tue Nov 17 08:50:30 +0000 2015", "geo": {"type": "Point", "coordinates": [14.5816011, 121.0492706]}}

161. {"original_text": "Still standing. Still dancing. \u2b50\ufe0f #hope #love #art #words #street #atnight #paris @ Le Marais - Paris\u2026 https://t.co/RwyvB7HVPO", "manual_label": "1", "hashtags": ["hope", "love", "art", "words", "street", "atnight", "paris"], "filtered_text": "still standing still dancing le marai paris", "original_id": 666751814117625857, "created_time": "Tue Nov 17 22:56:34 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8599619, 2.3565491]}}

162. {"original_text": "Moulin Rouge and absinthe at midnight #moulinrouge #paris #absinthe #midnightinparis @ Moulin Rouge https://t.co/CxWZNxmZrW", "manual_label": "1", "hashtags": ["moulinrouge", "paris", "absinthe", "midnightinparis"], "filtered_text": "moulin roug absinth midnight moulin roug",

"original_id": 666752805051371520, "created_time": "Tue Nov 17 23:00:31 +0000 2015", "geo": {"type": "Point", "coordinates": [48.88416667, 2.3325]}}

163. {"original_text": "#TWU is praying for Paris.\n#PrayForParis #GodsNotDead @ Margo Jones Auditorium Texas Women's University https://t.co/hI7h20p5nl", "manual_label": "3", "hashtags": ["TWU", "PrayForParis", "GodsNotDead"], "filtered_text": "pray paris margo jone auditorium texa women univers", "original_id": 666835170494980096, "created_time": "Wed Nov 18 04:27:48 +0000 2015", "geo": {"type": "Point", "coordinates": [33.22373464, -97.12973263]}}

164. {"original_text": "Living in the same arrondissement as where the #paris attacks occurred, I have not wanted to post,\u2026 https://t.co/jVEaVN2AAL", "manual_label": "2", "hashtags": ["paris"], "filtered_text": "live arrondiss attack occurred i want post", "original_id": 666709434501103616, "created_time": "Tue Nov 17 20:08:10 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86746503, 2.36418438]}}

165. {"original_text": "@AntBit Great call! I was thinking Waits, but of course those three guys. I saw this in open air cinema in Paris. #Paris", "manual_label": "6", "hashtags": ["Paris"], "filtered_text": "great call i think waits cours three guys i saw open air cinema paris", "original_id": 666703890537160704, "created_time": "Tue Nov 17 19:46:08 +0000 2015", "geo": {"type": "Point", "coordinates": [51.42872, -0.34215]}}

166. {"original_text": "C'est beau ! #liberteegalitefraternite #paris #france #stadedefrance #ruedecharonne #lepetitcambodge\u2026 https://t.co/azKTfHS2dx", "manual_label": "1", "hashtags": ["liberteegalitefraternite", "paris", "france", "stadedefrance", "ruedecharonne", "lepetitcambodge"], "filtered_text": "c est beau", "original_id": 666733238635876352, "created_time": "Tue Nov 17 21:42:46 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}

167. {"original_text": "#CCUFB\nInside the Warmup\nLiberty Week!! \nATTACK\n#1TEAM\n#BAM\n#builtCHANTtough\n#rarebreed @ Coastal\u2026 https://t.co/NHqmRl1E5l", "manual_label": "6", "hashtags": ["CCUFB", "1TEAM", "BAM", "builtCHANTtough", "rarebreed"], "filtered_text": "insid warmup liberti week attack coastal", "original_id": 666758712632197120, "created_time": "Tue Nov 17 23:23:59 +0000 2015", "geo": {"type": "Point", "coordinates": [33.79887401, -79.01552429]}}

168. {"original_text": "First espresso \u2615\ufe0f. First heart attack \ud83d\udc94. #enoughenergytorunfrommiamitochicago #jkhaterunning @\u2026 https://t.co/Zns5dlMKKc", "manual_label": "1", "hashtags": ["enoughenergytorunfrommiamitochicago", "jkhaterunning"], "filtered_text": "first espresso first heart attack o g c", "original_id": 666755944785686528,

"created_time": "Tue Nov 17 23:12:59 +0000 2015", "geo": {"type": "Point", "coordinates": [25.7731972, -80.1900635]}}

169. {"original_text": "Forrester Research #Sales #Job: Sales Development Specialist (#Paris) https://t.co/NQGlNsGeb6 #Jobs #Hiring", "manual_label": "5", "hashtags": ["Sales", "Job", "Paris", "Jobs", "Hiring"], "filtered_text": "forrest research sale develop specialist", "original_id": 666716002776805376, "created_time": "Tue Nov 17 20:34:16 +0000 2015", "geo": {"type": "Point", "coordinates": [48.856614, 2.3522219]}}

170. {"original_text": "#hnytwtr id65,BR,{1}WEB_X,POST attack from 186.202.153.x", "manual_label": "6", "hashtags": ["hnytwtr"], "filtered_text": "id65 br 1 web_x post attack 186 202 153 x", "original_id": 666693491490234368, "created_time": "Tue Nov 17 19:04:49 +0000 2015", "geo": {"type": "Point", "coordinates": [-10.0, -55.0]}}

171. {"original_text": "My shadow on a French street! #paris #streetstyle #parisstreet #parisnights # @ Paris, France https://t.co/bCK4p3eASR", "manual_label": "1", "hashtags": ["paris", "streetstyle", "parisstreet", "parisnights"], "filtered_text": "my shadow french street paris franc", "original_id": 666792664805847041, "created_time": "Wed Nov 18 01:38:54 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

172. {"original_text": "#TousAuBistrot , mais #bouare d'abord ! (@ CopperBay in Paris, \u00cele-de-France) https://t.co/NGNpwpPAnk", "manual_label": "1", "hashtags": ["TousAuBistrot", "bouare"], "filtered_text": "mai d abord copperbay paris le de france", "original_id": 666690099971792897, "created_time": "Tue Nov 17 18:51:20 +0000 2015", "geo": {"type": "Point", "coordinates": [48.869764, 2.357482]}}

173. {"original_text": "Can you find Paris on the map? Just try it at https://t.co/t8uzWiaOqr #Paris", "manual_label": "1", "hashtags": ["Paris"], "filtered_text": "can find pari map just tri", "original_id": 666670916181602304, "created_time": "Tue Nov 17 17:35:07 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.35099]}}

174. {"original_text": "#Paris #Freedom #DefeatISIS #UniteAgainstHatred #ToleranceWins https://t.co/RfLyh5Etb3", "manual_label": "2", "hashtags": ["Paris", "Freedom", "DefeatISIS", "UniteAgainstHatred", "ToleranceWins"], "filtered_text": "", "original_id": 666693386431238144, "created_time": "Tue Nov 17 19:04:24 +0000 2015", "geo": {"type": "Point", "coordinates": [67.64943718, 24.92204052]}}

175. {"original_text": "#Stigmabase | DE - Wie mit unseren Kindern \u00fcber den Terror in Paris sprechen? \u00a0-\u00a0Die Bilder und Filmaufnahmen von\u2026 https://t.co/pMSAuRDSZQ", "manual_label": "6", "hashtags": ["Stigmabase"], "filtered_text": "de wie mit unseren kindern ber den terror pari sprechen die bilder und filmaufnahmen von", "original_id": 666678106271256577, "created_time": "Tue

Nov 17 18:03:41 +0000 2015", "geo": {"type": "Point", "coordinates": [52.52786948, 13.42912598]}}

176. {"original_text": "I\u2764\ufe0fParis. Paris hearts. Graffiti on Parisian walls after the Paris Attacks on Friday. #parisheart\u2026 https://t.co/IUYJSdJ06Q", "manual_label": "3", "hashtags": ["parisheart"], "filtered_text": "i paris pari hearts graffiti parisian wall pari attack friday", "original_id": 666599996536541184, "created_time": "Tue Nov 17 12:53:18 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

177. {"original_text": "#autumn in #paris #jardin #albertkhan #boulognebillancourt #gardens #zen\u2026 https://t.co/AKluemrdsh", "manual_label": "1", "hashtags": ["autumn", "paris", "jardin", "albertkhan", "boulognebillancourt", "gardens", "zen"], "filtered_text": "", "original_id": 666689182421688321, "created_time": "Tue Nov 17 18:47:42 +0000 2015", "geo": {"type": "Point", "coordinates": [48.84211043, 2.22745657]}}

178. {"original_text": "\"for Paris is a moveable feast\" #parisstreetstyle #paris @ Paris, France https://t.co/vmEXLudwz8", "manual_label": "1", "hashtags": ["parisstreetstyle", "paris"], "filtered_text": "for pari moveabl feast paris franc", "original_id": 666760841421651968, "created_time": "Tue Nov 17 23:32:27 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

179. {"original_text": ". @GregPalkot And @GOP wonders why America prefers hearing @realDonaldTrump say \"I'll bomb the sh!t out of them\". #ISIS", "manual_label": "2", "hashtags": ["ISIS"], "filtered_text": "and wonder america prefer hear say i ll bomb sh t them", "original_id": 666602308361445376, "created_time": "Tue Nov 17 13:02:29 +0000 2015", "geo": {"type": "Point", "coordinates": [43.2314637, -75.44843506]}}

180. {"original_text": "The tweet with the most impact of the #exrdi Trend, was published by @s_bombero: https://t.co/lrkYxyXbkt (15 RTs) #trndnl", "manual_label": "6", "hashtags": ["exrdi", "trndnl"], "filtered_text": "the tweet impact trend publish 15 rts", "original_id": 666680502187782144, "created_time": "Tue Nov 17 18:13:12 +0000 2015", "geo": {"type": "Point", "coordinates": [56.9547, -98.309]}}

181. {"original_text": "Still standing \ud83c\uddeb\ud83c\uddf7 #fluctuatnecmergitur #eiffeltower #jesuisparis #paris #paris7 #france @ Tour Eiffel https://t.co/1r7dS8KCQl", "manual_label": "1", "hashtags": ["fluctuatnecmergitur", "eiffeltower", "jesuisparis", "paris", "paris7", "france"], "filtered_text": "still stand u e ce tour eiff ql", "original_id": 666678914303987712, "created_time": "Tue Nov 17 18:06:54 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}

182. {"original_text": "#paris #eiffeltower en bleu blanc rouge @ Tour Eiffel https://t.co/pvaywXO3CZ", "manual_label": "1", "hashtags": ["paris", "eiffeltower"], "filtered_text": "en bleu blanc roug tour eiffel", "original_id": 666543853797683200,

"created_time": "Tue Nov 17 09:10:13 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}

183. {"original_text": "Physician (Dermatologist) - Department of Veterans Affairs: (#Lebanon, PA) https://t.co/8jRmLNO6n7 #Physician #Veterans #Job #Jobs #Hiring", "manual_label": "5", "hashtags": ["Lebanon", "Physician", "Veterans", "Job", "Jobs", "Hiring"], "filtered_text": "physician dermatologist depart veteran affairs pa", "original_id": 666725964194312192, "created_time": "Tue Nov 17 21:13:51 +0000 2015", "geo": {"type": "Point", "coordinates": [40.3409251, -76.4113497]}}

184. {"original_text": "After Paris Attacks, CIA Director Rekindles Debate Over Surveillance - New York Times https://t.co/akG4FSEmgP #news https://t.co/0hJ228enTT", "manual_label": "2", "hashtags": ["news"], "filtered_text": "after pari attacks cia director rekindl debat over surveil new york time https t co 0hj228entt", "original_id": 666499528300752898, "created_time": "Tue Nov 17 06:14:05 +0000 2015", "geo": {"type": "Point", "coordinates": [36.06801502, -79.79721587]}}

185. {"original_text": "Silent tributes as Kendal joins Europe-wide silence for Paris terror attacks https://t.co/rP0ciWCJ9W #Cumbria https://t.co/J5RUmWNNwl", "manual_label": "2", "hashtags": ["Cumbria"], "filtered_text": "silent tribut kendal join europe wid silenc pari terror attack https t co j5rumwnnwl", "original_id": 666527047972335617, "created_time": "Tue Nov 17 08:03:26 +0000 2015", "geo": {"type": "Point", "coordinates": [54.89453982, -2.93380223]}}

186. {"original_text": "#EnigmaticSmile full length album \u266b Panic Attack by Annisokay (at Horney Get Store) \u2014 https://t.co/jaMWHRKGoY", "manual_label": "2", "hashtags": ["EnigmaticSmile"], "filtered_text": "full length album panic attack annisokay at horney get store", "original_id": 666829721049546752, "created_time": "Wed Nov 18 04:06:09 +0000 2015", "geo": {"type": "Point", "coordinates": [-8.35855, 114.64388]}}

187. {"original_text": "#MARDICESTBISTROT \ud83d\ude4f\ud83c\udffb\n\n#paris #frenchrestaurant #bistrot #traditional #love #prayforparis #cheffe\u2026 https://t.co/wFZSg7u6Qn", "manual_label": "3", "hashtags": ["MARDICESTBISTROT", "paris", "frenchrestaurant", "bistrot", "traditional", "love", "prayforparis", "cheffe"], "filtered_text": "n o v fe qn", "original_id": 666701546831405057, "created_time": "Tue Nov 17 19:36:50 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86472222, 2.33222222]}}

188. {"original_text": "#paris #pont @ Pont Alexandre III https://t.co/2ynt3ytZ5i", "manual_label": "6", "hashtags": ["paris", "pont"], "filtered_text": "pont alexandr iii", "original_id": 666541640807387137, "created_time": "Tue Nov 17 09:01:25 +0000 2015", "geo": {"type": "Point", "coordinates": [48.86361111, 2.31361111]}}

189. {"original_text": "#catacombs #paris @ Catacombs of Paris https://t.co/M9ZDymxHL4", "manual_label": "1", "hashtags": ["catacombs", "paris"],

"filtered_text": "catacomb pari", "original_id": 666741938641616896, "created_time": "Tue Nov 17 22:17:20 +0000 2015", "geo": {"type": "Point", "coordinates": [48.83400833, 2.33232222]}}

190. {"original_text": "#Paris tonight outside Le Bataclan nightclub. Live reports on #ParisAttacks 4/5/6/11p nbcphiladelphia\u2026 https://t.co/J994jFzgRt", "manual_label": "2", "hashtags": ["Paris", "ParisAttacks"], "filtered_text": "tonight outsid le bataclan nightclub live report 4 5 6 11p nbcphiladelphia", "original_id": 666708589000720384, "created_time": "Tue Nov 17 20:04:49 +0000 2015", "geo": {"type": "Point", "coordinates": [48.863121, 2.3708701]}}

191. {"original_text": "Good morning, Paris! :) #Sunrise 08:02, noon 12:36, sunset 17:09 CET (UTC+1), November 17. Day length: 9h 7m.", "manual_label": "4", "hashtags": ["Sunrise"], "filtered_text": "good morning paris 08 02 noon 12 36 sunset 17 09 cet utc 1 novemb 17 day length 9h 7m", "original_id": 666511638347055104, "created_time": "Tue Nov 17 07:02:12 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.351]}}

192. {"original_text": "#chezmatante #creditmunicipal #art #expo #basket #inParis #villelumiere @ Cr\u00e9dit Municipal de Paris https://t.co/ralWjqNbBC", "manual_label": "1", "hashtags": ["chezmatante", "creditmunicipal", "art", "expo", "basket", "inParis", "villelumiere"], "filtered_text": "cr dit municip de pari", "original_id": 666810269134225409, "created_time": "Wed Nov 18 02:48:51 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8594704, 2.35761]}}

193. {"original_text": "Lovers #concorde @ Paris, France https://t.co/1MmI30flNt", "manual_label": "1", "hashtags": ["concorde"], "filtered_text": "lover paris franc", "original_id": 666670618864128000, "created_time": "Tue Nov 17 17:33:56 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

194. {"original_text": "Caf\u00e9 Arabe #buen\u00edsimo en Bomb\u00f3n Oriental @ Santiago, Chile https://t.co/G5AokqCwwM", "manual_label": "1", "hashtags": ["buen\u00edsimo"], "filtered_text": "caf arab en bomb n orient santiago chile", "original_id": 666729620105003008, "created_time": "Tue Nov 17 21:28:23 +0000 2015", "geo": {"type": "Point", "coordinates": [-33.45, -70.6667]}}

195. {"original_text": "Wonder if the french want their statue back. #paris @ Starlite Hotel https://t.co/8LmRs7mvSo", "manual_label": "1", "hashtags": ["paris"], "filtered_text": "wonder french want statu back starlit hotel", "original_id": 666816107269984256, "created_time": "Wed Nov 18 03:12:03 +0000 2015", "geo": {"type": "Point", "coordinates": [42.124499, -80.105016]}}

196. {"original_text": "Thousands gather at oldmainstagram #paris #peace #remembering @ Penn State https://t.co/Z0mEuWGV0q", "manual_label": "3", "hashtags": ["paris", "peace", "remembering"], "filtered_text": "thousand gather oldmainstagram penn

state", "original_id": 666777959567527936, "created_time": "Wed Nov 18 00:40:28 +0000 2015", "geo": {"type": "Point", "coordinates": [40.7961816, -77.8593151]}}

197. {"original_text": "We Love our Country, We love our City and we are so Proud of it! From Como with \u2764\ufe0f #Paris #France\u2026 https://t.co/zlk7njXrN5", "manual_label": "1", "hashtags": ["Paris", "France"], "filtered_text": "we love country we love citi proud it from como", "original_id": 666777638640291840, "created_time": "Wed Nov 18 00:39:11 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}

198. {"original_text": "#Ir\u00e1n asegura que se deber\u00eda juzgar a Estados Unidos por haber creado el Estado Isl\u00e1mico https://t.co/TBfg968F0S https://t.co/eg3onfrQVY", "manual_label": "2", "hashtags": ["Ir\u00e1n"], "filtered_text": "asegura que se deber a juzgar estado unido por haber creado el estado isl mico https t co eg3onfrqvi", "original_id": 666553186602369024, "created_time": "Tue Nov 17 09:47:18 +0000 2015", "geo": {"type": "Point", "coordinates": [-16.4963067, -68.1164723]}}

199. {"original_text": "\ud83d\udcf7\ud83d\udc49 oneguyinthecity \n#fallfashion #fallvibes @ Paris, France https://t.co/3XY7I837aa", "manual_label": "1", "hashtags": ["fallfashion", "fallvibes"], "filtered_text": "oneguyinthec o es paris franc aa", "original_id": 666667787201282048, "created_time": "Tue Nov 17 17:22:41 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8567, 2.3508]}}

200. {"original_text": "This photo was taken a while back...when I visited #Paris I saw diversity and love - all the deaths\u2026 https://t.co/flLVkKYmOY", "manual_label": "3", "hashtags": ["Paris"], "filtered_text": "thi photo taken back when i visit i saw divers love deaths", "original_id": 666507400032223232, "created_time": "Tue Nov 17 06:45:21 +0000 2015", "geo": {"type": "Point", "coordinates": [48.8590256, 2.29811714]}}