Twitter Floods when it Rains:
A Case Study of the UK Floods in early 2014

Antonia Saravanou
University of Athens
antoniasar@di.uoa.gr

Dimitrios Gunopulos
University of Athens
dg@di.uoa.gr

George Valkanas
Stevens Institute of Technology
gvalkana@stevens.edu

Gennady Andrienko
Fraunhofer Institute IAIS, DE
gennady.andrienko@iais.fraunhofer.de

ABSTRACT

Twitter is one of the most prominent social media platforms nowadays. A primary reason that has brought the medium at the spotlight of academic attention is its real-time nature, with people constantly uploading information regarding their surroundings. This trait, coupled with the service's data access policy for researchers and developers, has allowed the community to explore Twitter's potential as a news reporting tool. Finding out promptly about newsworthy events can prove extremely useful in crisis management situations. In this paper, we explore the use of Twitter as a mechanism used in disaster relief, and consequently in public safety. In particular, we perform a case study on the floods that occurred in the United Kingdom during January 2014, and how these were reflected on Twitter, according to tweets (i.e., posts) submitted by the users. We present a systematic algorithmic analysis of tweets collected with respect to our use case scenario, supplemented by visual analytic tools. Our objective is to identify meaningful and effective ways to take advantage of the wealth of Twitter data in crisis management, and we report on the findings of our analysis.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

Keywords

microblogs; live web; event detection; floods; visualization

1. INTRODUCTION

Microblogging platforms are at the core of what is known as the *Live Web*: the most dynamic, and fast changing portion of the web, where content is generated *constantly* by the *users*, in snippets of information. Therefore, the *Live Web* is a good data source for event detection and tracking [11, 12, 15, 17, 19], because it reflects what is happening in the physical world in a timely manner.

The gravity of events discussed in social media can vary significantly, ranging from leisure activities and hobbies (e.g., sports, movies, music, etc) to more practical cases, affecting day-to-day activities (e.g., construction sites, road blocks) to extreme situations (e.g., floods, earthquakes). This variety is a direct outcome of the diversity of the participating users and their interests. Needless to say that identifying, and subsequently tracking, events of the latter form (i.e., extreme situations) is exceedingly important. In such cases, the responsible authorities need access to high-quality, accurate information to appropriately allocate resources, assign tasks and coordinate with one another. Social media platforms can fill in the gap of the data provider, particularly in cases where there is no existing infrastructure.

Unfortunately, (timely) event detection is a daunting task, even in more conventional domains [2, 9]. To further complicate matters, social media impose additional constraints: high volumes of data in various formats, which is generally noisy, and thematically and linguistically diverse. Extracting the exact location of an event is also not trivial, because the volume of GPS-accurate information is low [14]. For this reason, researchers have proposed techniques to address the sparsity of spatial information [4, 8, 16]. One should also take into account that social media users may discuss events that are far beyond their physical surroundings, but are of interest to them nevertheless.

The characteristics described above can be found - in various proportions - in all social media platforms. Twitter [1], one of the most prominent social media nowadays, is no exception. The platform received widespread public attention after the protests following the Iranian election (2009-2010) and Egyptian Revolution (2011), where the service played a major role. It has also been very well received by researchers and developers, due to its data access policy. More specifically, the service grants free access [2] to 1% of all public tweets in a streaming fashion – which facilitates real-time event detection – as well as to all types of data it has, albeit with certain restrictions, in a query-based fashion.

Taking into account our previous discussion, in this paper, we are interested in utilizing social media data for *social good*. More specifically, we want to explore the potential of social media information in disaster management and relief, and provide meaningful and effective techniques towards this goal. Given the generality of our objective and the plethora of available techniques one could employ, we will limit ourselves to a case study related to disaster management. Within this scope, we will attempt to answer the following research questions:

---

[1]https://twitter.com/

[2]For non-commercial use only

Q1: How can we identify the areas that have been hit the most by an event? Identifying heavily hit areas is a key step in disaster relief. Knowing *where* to dispatch emergency response units is an indispensable piece of information, without which we are unable to provide any help.

Q2: How effective can we be in identifying these areas? Considering that in disaster management, emergency units are typically confronted with life threatening situations, it is imperative that they have robust and effective techniques to base their decisions on. Coupled with the fact that our primary data source is noisy, it is important to know *in advance* their effectiveness limitations. In principle, we need to evaluate and validate the results that we obtain, preferably in an offline setting, before applying them in real life.

Q3: Can we identify areas that have been striken by the event in a similar manner? It would be useful to identify areas where the event manifested itself in similar ways. In such cases, emergency response units would be able to apply / transfer techniques that performed well in other regions. Secondly, but equally important, a similar manifestation of an event may be the result of an underlying connection between the two regions that the authorities may need to consider.

Our use case that will help us answer the previous questions is the floods that took place in the United Kingdom during early 2014. We will utilize data harvested from Twitter, focusing on tweets (i.e., posts) that discuss this particular event. By zooming in on such tweets, we are able to monitor the evolution of the event, as it was experienced and portrayed by Twitter users over a 5 day period. We subsequently apply data mining techniques, following a rigorous methodology, aided by appropriate visualization tools. We also evaluate our results using ground truth information provided by external, independent sources, thereby verifying our approach. This will also allow us to make the most sense out of our collected data and suggest meaningful ways to process such information in the case of similar events in the future. Finally, we discuss lessons learned from our engagement in this particular project.

In short, we make the following contributions in this paper:

a) We perform a case study, utilizing social media data (i.e., tweets) over a 5 day period, regarding the UK floods that took place in early 2014.

b) We present in detail a rigorous processing methodology, to reach meaningful and tangible results, and to answer our research questions.

c) We use independently provided ground truth information regarding the same period, to allow us to verify our observations and findings.

The rest of the paper is organized as follows: Section 2 briefly presents related work. Section 3 discusses data collection and cleaning, followed by Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORK

The nature of social media data has led to an increased use of the medium for the purposes of event identification. Early works focus on events of specific types e.g. news [13] or political debates and elections. This is typically done by either whitelisting users (e.g., news reporting agencies) or by building on the premise that the topic is inherently polarized, in which case we can monitor the reactions of user communities. Emergency events, such as the ones we consider here (i.e., floods, fires), are not polarized, rendering such techniques useless. On the other hand, whitelisting will only provide access to the raw data – which we can retrieve through other means as well –, without offering any additional insights.

The most characteristic (and successful) example of event identification through social media information is [12]. This work focuses on earthquakes and its main objective is to accurately extract the location of an earthquake. The authors rely on a manually crafted lexicon to match geotagged tweets related to the event. They propose a model that incorporates established scientific theories on wave propagation, through which they identify the origin of an earthquake. These particular theories, however, do not apply to floods, rendering these techniques inapplicable for our use case.

Type-independent event identification has also been a hot research area recently, to avoid the manual or semi-automatic generation of lexicons [2, 19, 17]. These techniques are more complex than the one we propose in this paper, but their objective is in identification and much less in understanding the spread of information on social media or how an event is portrayed online. Although event identification is clearly the first step to disaster management and relief, we believe that our current analysis may provide additional insights on how to utilize social media information.

Recent research attempts have also put social media to the test for flooding events [7, 3]. Such attempts are usually constrained to a statistical analysis of the collected dataset, e.g., #tweets, #users, hashtag distribution, and secondarily to an analysis related to the flooding event itself, e.g., vicinity of tweets to the event. Even in these cases, the evaluation implies that we have access to high quality sensing devices. Although this does not invalidate the findings of the research works, in practice, it would render social media useless for the purposes of flood disaster management.

The most closely related work regarding flooding events is the one in [5]. The authors use their framework to visually examine spatio-temporal data regarding the floods that took place in Germany in the summer 2013. They focus on streamlining a visual analytics workflow that will assist in detecting significant events, and discuss their hypothesis and findings. On the contrary, in our current work, visualization tools are supplementary, as we focus on a methodology and accompanying techniques that will help us identify heavily stricken areas.

3. DATA COLLECTION AND CLEANING

We will start by describing the data we collected to perform our case study. Posts on Twitter, generally known as *tweets*, are short snippets of text, up to 140 characters. Aside from text, they may contain features like (shortened) *hyperlinks* to external sources, *hashtags*, i.e., author-generated tags describe the topic of a tweet (e.g. #sports), *mentions* of other Twitter users, or other data formats, including images and videos.

We used our custom built crawler [18] to collect tweets from the Twitter service, using the Streaming API. The Streaming API returns a sample of all public tweets, as these are posted in the service, granting us access to up-to-date information. Given our interest in the floods that took place in the United Kingdom (UK), we limited ourselves to tweets posted from this particular area. To achieve this, we used a bounding box that covers the entire UK, namely $[(-13.4139, 49.1621), (1.7690, 60.8547)]$ [3], and used that as a first filter for collecting the data. This way, the sampling policy is applied directly to those posts that, according to the service, have originated from a location within the requested bounds. In

---

[3]Coordinates in GPS format, obtained from `https://www.flickr.com/places/info/23424975`

some cases, we may receive a tweet that does not have an associated GPS coordinate, because the service has concluded that it falls within the requested bounding box through other means, e.g. the user's profile. We do not consider such tweets during our analysis, because our approach is based on the specific geo-olocation of a tweet. More information about the bounding box filter is available at `https://dev.twitter.com/streaming/overview/request-parameters`.

We applied the data harvesting technique described above for a 5 day period, during early 2014. Once a tweet is received, we store all of its information externally for offline processing. Table 1 summarizes some basic statistics of the collected data. The entire dataset consists of more than 2.3 million geotagged tweets. The first and last days contain about half of the tweets compared to the rest, as our crawl only covered half a day on each of these occasions. The late start was due to starting the crawler as soon as we were informed of the event. We ended the crawl once we saw that the flooding incident had subsided.

Table 1: Dataset Statistics

Period	Total #Tweets	#Flood Related Tweets
Jan. 13	351140	2728
Jan. 14	577151	4973
Jan. 15	569108	4159
Jan. 16	578553	4994
Jan. 17	275358	3490
Total	2351310	20344

Extracting flood related tweets. A problem with our original dataset is that it contains tweets that have been filtered only on the grounds of location. However, because of the high diversity of Twitter users and their interests, it is likely that many tweets will be completely unrelated to the event which we are monitoring. For this reason, we need to perform an additional filtering step, to keep only those posts that are related to floods.

To achieve this, we built a custom lexicon, containing tokens related to our event. Note that this is the norm for monitoring events of a particular type in Twitter [11, 13, 12]. We started with a very small *seed set* of 13 related tokens, including "*rain, flood, weather, storm, showers*", etc. We parsed the entire dataset and extracted all of the tokens – excluding mentions, i.e., @*username* – that contain at least one of the seed set keywords as a substring.

Following the methodology described above, we obtained a lexicon with 1546 distinct keywords, much larger in size than the seed set. This lexicon contains the seed keywords in various forms, e.g. *raining*, *floods*, *#ukweather*, etc. However, our approach also yielded a lot of false positives, e.g., *brain*, *train*, etc. We carefully reviewed each keyword and discarded everything that is not related to our use case. The final, cleaned version of our lexicon contains 456 distinct keywords. Although this is a laborious process, it only needs to be performed once.

Table 2 contains the top-10 keywords for the two lexicons: the one that was compiled directly from the collected tweets, and the cleaned version, containing only the keywords related to our event. It is important to note that keywords with a high number of occurrences (as seen from Table 2) are generic enough and are related to the event *type*, not to the particular location. This means that we can use these keywords for the same purpose (flood monitoring) in other locations as well. We make our flood lexicon available to the community for reviewing and use [4].

Table 2: Lexicon Keywords

	Original Lexicon (1546 keywords)		Flood Lexicon (456 keywords)	
Rank	**Keyword**	**In #Tweets**	**Keyword**	**In #Tweets**
1	*rain*	11235	*rain*	11235
2	*train*	6499	*weather*	3331
3	*training*	4593	*snow*	1006
4	*weather*	3331	*raining*	997
5	*brain*	1747	*rainbow*	419
6	*trains*	1251	*storm*	333
7	*snow*	1006	*showers*	273
8	*raining*	997	*rainy*	249
9	*trainers*	813	*flooding*	215
10	*drained*	435	*flooded*	214

Having our flood-related lexicon in place, we iterate over the entire collection for a second time. During this step, we select tweets as follows: We tokenize the text of the tweets and keep only those that contain an *exact* match with at least one of the keywords of our flood lexicon. Algorithm 1 presents the pseudocode for selecting the flood related tweets. Given that, at this point, our lexicon contains only flood-related tokens, we expect the majority of the extracted tweets to be discussing the event of our use case, making them safe to use in our subsequent analysis. We say "the majority" because some lexicon keywords may have a slightly altered meaning on occasions. For instance, the expression "*be under the weather*" is most likely unrelated to a flooding event, although it contains the seed keyword "weather". To that end, the authors did a preliminary manual analysis of 1000 randomly selected tweets, that contain at least one keyword from the flood lexicon. This analysis showed that 88.5% of the sample is relevant to the event that we care about.

Algorithm 1 `Selection of Flood-related Tweets`

Input: Set of tweets $\mathcal{T}$, Lexicon $\mathcal{L}$
Output: Set of flood-related tweets $\mathcal{T}_{flood}$
1: $\mathcal{T}_{flood} \leftarrow \emptyset$;
2: **for** every tweet $t \in \mathcal{T}$ **do**
3:     tokens $\leftarrow$ split $t.text$ into tokens;
4:     **for** every token $tkn \in tokens$ **do**
5:         **if** $\mathcal{L}$.contains(tkn) **then**
6:             Add t to $\mathcal{T}_{flood}$;
7:             **break**;
8: **return** $\mathcal{T}_{flood}$

---

4. EXPERIMENTS

As we have already pointed out, we are interested in techniques that will provide meaningful insights to our collected data. In particular, we have noted three specific questions that we would like to answer, through our subsequent experimental analysis:

Q1: How can we identify the areas that have been hit the most by an event?

Q2: How effective can we be in identifying these areas?

Q3: Can we identify areas that have been striken by the event in a similar manner?

4.1 Generating monitored regions

The major concern of question **Q1** is the identification of the areas that have been hit by the event the most. This type of knowledge
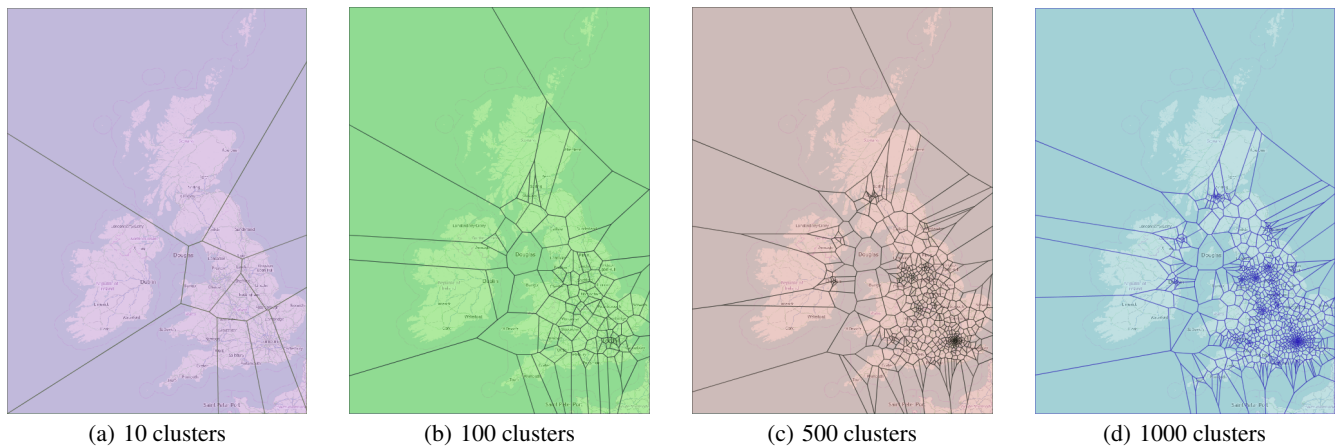
(a) 10 clusters   (b) 100 clusters   (c) 500 clusters   (d) 1000 clusters

Figure 1: Spatial clusters generated by k-Means clustering on the entire (2.3M) collection of tweets

is particularly important in disaster management and relief situations, as emergency units need to know *where* their assistance is required. Given that we need to identify *areas* or *regions*, we need to go beyond the GPS coordinates of a single tweet. Therefore, we need to aggregate the geotagged information of our tweets to form these larger areas. For this purpose, we apply k-Means clustering, which will create k non-overlapping regions, as follows.

We extract the geotagging information from all of the tweets of the *original* dataset. This gives us a little more than 2.3M GPS coordinates in (longitude, latitude) form, covering the entire country (UK). We use the original dataset because it contains a lot more tweets, thus we expect the clusters to better reflect the actual underlying population. The idea is to perform spatial clustering using the GPS coordinates of the tweets, to obtain larger regions where more than a few tweets will be posted from. Before the clustering step, we convert the GPS coordinates to Cartesian ones, using Mercatorian map projections [10]. This step is necessary for k-Means to work properly with the L_2 (Euclidean) distance which we used.

A well-known drawback of k-Means is the selection of an appropriate k value for the number of clusters. For this reason, we experimented with different k values ($k = 10, 100, 500, 1000$) and report on the results for each case. Although our selection of k values is far from an exhaustive search, other clustering algorithms, e.g., DBScan [6], require more parameters to tune, making the search space much more difficult to explore and interpret. Alternative approaches, such as Agglomerative Hierarchical Clustering [6], would also be difficult to evaluate empirically (see **Q2**), and would still require manual effort in selecting the clusters to evaluate. We acknowledge this limitation of our current study, and leave as future work the study of alternative clustering techniques.

Figure 1a)-d) shows the boundaries of each generated cluster as Voronoi polygons, for the different values of k, overlayed on top of a UK map [5]. We can clearly see that as we increase the number of clusters k, we get more splits in the densely populated areas (London, Manchester, Leeds, Glasgow, etc), as opposed to more rural areas. As a result, our current analysis could be seen as a weak proxy for hierarchical clustering, cutting the hierarchical dendrogram at different levels (heights).

4.2 Identifying flood-affected areas

In the previous section, we showed how to generate large regions, starting from simple geotagged tweets. We now propose

---

[5]Visualization through Visual Analytics tool [1]. Ireland is included due to the original bounding box that was used.

how to identify the areas that were mostly affected by the floods, thereby answering **Q1**. Towards this goal, we will prioritize the generated regions by their potential of being affected by the flood, and propose 3 schemes to do this. The scheme would practically dictate the order with which emergency response units should attend to each region.

Given that the number of regions to check is quite high (1000 regions at most), the prioritization scheme will also allow us to reduce the evaluation cost. A good prioritization scheme should return highly affected areas first, and less affected areas afterwards. We can then evaluate the top-n regions, as returned by each scheme.

The prioritization approaches one may use for this purpose are:

- By **#tweets**: This is the simplest scheme and serves as our baseline. The regions are ordered by the number of tweets posted from it, regardless of being flood-related or not.

- By **#flood-related tweets**: The regions are ordered by the number of flood-related tweets posted from that area.

- By **SNR**: A potential problem of the previous scheme is that densely populated areas are more likely to have a high number of flood related tweets. We counter this via a Signal-to-Noise Ratio approach. Each region r is assigned a score

$$score(r) = \frac{\#\texttt{flood-related tweets in r}}{\#\texttt{tweets in r}}$$

and areas are ordered by their respective score.

4.3 Evaluation

To evaluate our results, we use two independent sources as ground truth information. The first source is the monthly hydrological report for January 2014, published by the Centre for Ecology & Hydrology, of the Natural Environment Research Council in the UK, available at `http://www.ceh.ac.uk/data/nrfa/nhmp/hs/pdf/HS_201401.pdf`. The second source was published by UK's National Weather Service, Met Office, and discusses the storms that hit the UK during the period we monitored. That report is available at `http://www.metoffice.gov.uk/media/pdf/n/i/Recent_Storms_Briefing_Final_07023.pdf`.

The evaluation took place as follows. From the k clusters obtained through k-Means clustering, we select the top-n (n=100 in our case), ordered by each scheme (#tweets, #flood tweets, SNR). We manually review each of the n areas, and compare it against the ground truth information that we have. We use a *Likert scale* ([1-5]) to specify the degree up to which an area has been affected
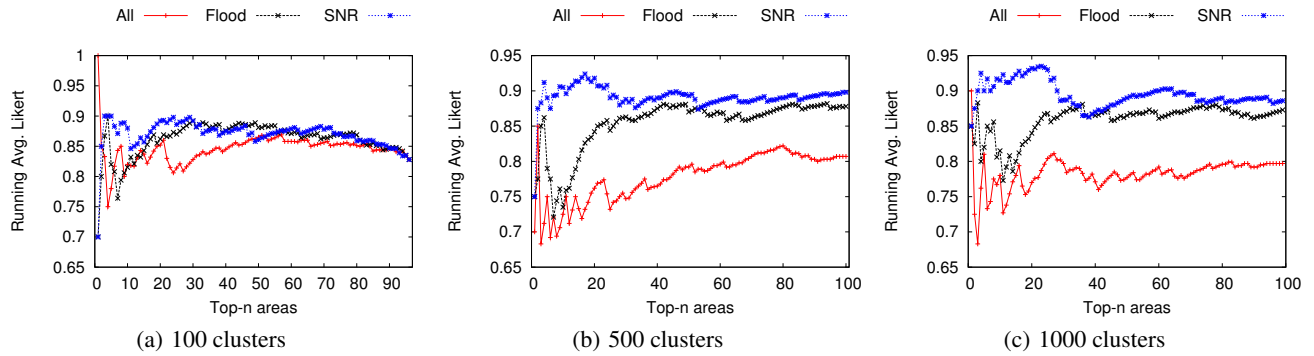
| (a) 100 clusters | (b) 500 clusters | (c) 1000 clusters |

Figure 2: Running average of the normalized Likert scores, of the top-n regions that were selected by each prioritization scheme

by flooding (1="not at all", 5="completely flooded"), and assign a score to each of the top-n areas. The assigned score for each area is normalized in the [0-1] range, to allow for easier comparison. We then compute and report the running average of the scores up to the i-th ranked area, using the formula

$$value_i = \frac{\sum_{j=1}^{i} \frac{likert\_score(j)}{5}}{i}$$

where $likert\_score(j)$ is the likert evaluation for the j-th region.

The result of this evaluation is given in Figure 2. On the x-axis we plot the top-n areas, whereas on the y-axis we plot the running average of percentages, using the above formula. We compare the 3 schemes: i) by number of tweets in the area (All) and ii) by number of flood-related tweets (Flood) and iii) by Signal-to-Noise Ratio (SNR).

We observe that for k=100 clusters, the schemes *Flood* and *SNR* behave almost the same, with SNR being slightly better at first. Both techniques perform better than the baseline approach (*All*), but their differences become blurred after the first 50 areas. Looking at Figure 1(b), we see that there is not much detail in the generated regions, even in areas like London, which might explain the minimal differences among the techniques.

When we increase the number of clusters to k=500 or k=1000, *Flood* and *SNR* clearly outperform the baseline. This shows that the number of social media users in an area is *not* a good proxy for the impact of an event. Counting the number of event-related posts (*Flood*) is a much better approach to identify areas that were hit harder. Nevertheless, Figures 2b)-c) lead us to believe that this technique shares some of the deficiencies of the baseline approach, due to the steep decline early on (around top-15 areas). Unlike the baseline, however, *Flood* improves much faster. On the other hand, *SNR* outperforms both of the other techniques, especially during the top-n areas. The scheme's performance, compared to that of *Flood*, is evidence of our argument that densely populated regions are more likely to have more event-related posts. Therefore, when we account for the number of users in that area, as *SNR* does, we manage to achieve even better performance. More specifically, *SNR* is able to maintain a running average of ~0.9 for the top-100 areas, never dropping below 0.85, providing us with quantitative data for question **Q2**.

Another thing to note here is that *Flood* exhibits the same decline during the first top-n regions, regardless of the number of clusters k. This leads us to conclude that this is a result of our collected data, rather than the prioritization scheme. The scale of our current use case does not allow us to dismiss this scheme as ineffective, however, it is clear that there are cases (such as this one) where this scheme may perform poorly. In emergency situations, where lives are at stake, decisions based on poor quality information can prove detrimental. We plan to experiment more with these schemes, and find out whether these properties hold in general.

4.4 Temporal Similarities of Affected Regions

One of our final objectives is to identify regions with similar behavior, in the way that the flooding event was perceived by Twitter users. The general implication of such a similarity is that there is an underlying connection, between these areas. This connection could be at the population level, e.g. the users have similar posting patterns when it comes to such events, or could be due to some other variable, e.g., a nearby river, or a problem in the plumbing system of those areas.

For this type of analysis, we start with our set of k clusters, obtained from k-Means. Each cluster is now a tuple and the output of this analysis will be groups of clusters. Because we are interested in temporal similarities of the regions, the features that we will use will be based on daily information.

We will consider the following sets of features, all of which are now applied to the flood related tweets.

- The **number of tweets** that were posted each day d from region r, denoted by $count_r^d$.

- The **ratio** of day d from region r is the fraction of tweets posted that day from that region, over all tweets posted from that region for the monitored period. Formally,

$$ratio_r^d = \frac{count_r^d}{\sum_{\forall d'} count_r^{d'}}$$

- **Speed** of day d is the difference between the ratio of day d and the preceding one.

$$speed_r^d = ratio_r^d - ratio_r^{d-1}$$

The rationale of this feature is to capture abrupt changes, and – hopefully – identify areas that were affected without sufficient notice.

We experimented with various combinations of these features, when performing the second level of clustering. Figures 3a-b) show two such groupings that were the result of clustering the regions using only the speed feature. The x-axis refers to the day i for which we evaluate $speed^i$, depicted on the y-axis. Each figure refers to a distinct 2nd level clustering, where we can clearly see the difference in the speeds of the regions. Figure 3c) visualizes the two clusterings on the map. Cluster 1 (in red) contains areas
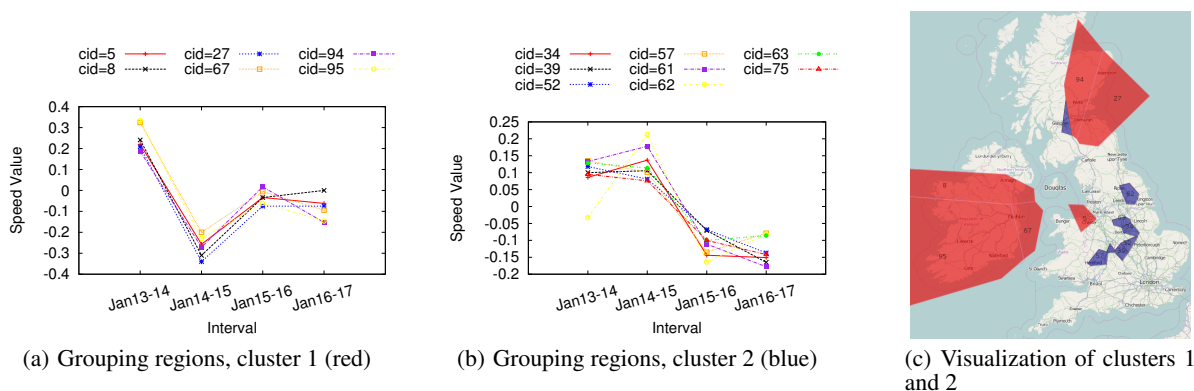
(a) Grouping regions, cluster 1 (red)　　(b) Grouping regions, cluster 2 (blue)　　(c) Visualization of clusters 1 and 2

Figure 3: Second level of clustering information, using the speed feature

from Scotland, Liverpool and Ireland, whereas Cluster 2 (in blue) contains areas mostly from the Midlands.

Figures 3a-b) illustrate the difference between areas from the two clusters. Areas in cluster 1 are mostly unaffected by floods, whereas cluster 2 contains regions with the opposite behavior. A difference in the trend of precipitation has also been verified with historical data from `http://www.weatheronline.co.uk/`. This is a characteristic example why the speed feature results in one of the best clusterings that we observed.

5. CONCLUSION & DISCUSSION

In this paper, we experimented with Twitter for the purposes of disaster management and relief. Using the floods that occurred in January 2014 in the United Kingdom as a use case, we collected geotagged tweets for analysis and insights. We proposed a methodology to clean the original dataset, build larger regions to monitor, and process the dataset to identify flood-stricken areas with high accuracy. We evaluated our findings against ground truth data, obtained from external, independent sources and reported on our findings.

Our work, however, is not without limitations. Our dataset is rather small (5 day period) and limited to a single use case. This begs the question of how generalizable are our results. We plan to collect more data on this particular event, and experiment with similar flooding events that have been reported in the past.

Another limitation is that our current analysis is restricted to a static analysis of the tweets. In our future work, we would like to develop online or streaming approaches, including clustering algorithms for our setting.

6. REFERENCES

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer-Verlag Berlin Heidelberg, 2013.

[2] H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In *WSDM*, 2010.

[3] M. A. Brovelli, G. Zamboni, C. A. Muñoz, and A. Bonetti. Exploring twitter georeferenced data related to flood events: an initial approach. 2014.

[4] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing. A latent variable model for geographic lexical variation. In *EMNLP*, 2010.

[5] G. Fuchs, N. Andrienko, G. Andrienko, S. Bothe, and H. Stange. Tracing the german centennial flood in the stream of tweets: First lessons learned. In *GEOCROWD '13*, 2013.

[6] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. 3rd edition, 2011.

[7] B. Herfort, J. P. de Albuquerque, S. Schelhorn, and A. Zipf. Exploring the geographical relations between social media and flood phenomena to improve situational awareness - A study about the river elbe flood in june 2013. In *AGIEL*, 2014.

[8] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsiouliklis. Discovering geographical topics in the twitter stream. In *WWW*, 2012.

[9] E. Keogh, S. Lonardi, and B. Y.-c. Chiu. Finding surprising patterns in a time series database in linear time and space. In *KDD*, 2002.

[10] D. H. Maling. *Coordinate Systems and Map Projections, 2nd Ed.* Pergamon Press, 1992.

[11] M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In *SIGMOD*, 2010.

[12] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW*, 2010.

[13] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *SIGSPATIAL-GIS*, 2009.

[14] A. Stefanidis, A. Crooks, and J. Radzikowski. Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 2013.

[15] J. Sutton, L. Palen, and I. Shlovski. Back-channels on the front lines: Emerging use of social media in the 2007 southern california wildfires. In *ISCRAM*, 2008.

[16] G. Valkanas and D. Gunopulos. Location extraction from social networks with commodity software and online data. In *ICDM Workshops (SSTDM)*, 2012.

[17] G. Valkanas and D. Gunopulos. How the live web feels about events. In *CIKM*, 2013.

[18] G. Valkanas, A. Saravanou, and D. Gunopulos. A faceted crawler for the twitter service. In *WISE 2014*.

[19] J. Weng and B.-S. Lee. Event detection in twitter. In *ICWSM*, 2011.