Analyzing the impact of natural disasters: A concept for spatio-temporal analyses of social media data

Jens Kersten*, Anna Kruspe*, Friederike Klan*

**German Aerospace Center (DLR)*
Institute of Data Science, Jena, Germany
<firstname>.<lastname>@dlr.de

***Abstract*—Twitter data is known to be a valuable source for rescue and helping activities in case of natural disasters and technical accidents in many countries worldwide. Several methods for tweet filtering and classification are available to analyze social media streams. However, rather than single tweets, analyzing the dynamics of reactions caused by disasters taking into account space and time is likely to reveal even more insights, e.g. regarding local event impacts on population and environment. In turn, this knowledge may help in developing more reliable methods for analyzing social media streams, e.g. for (sub-) event detection and tracking, that contribute to the improvement of disaster risk management strategies. As a starting point for this, suitable approaches for spatio-temporal analyses of Twitter data streams are discussed in this study. Besides the dimensions of time and space, the content of the tweets is also taken into account. Results from applying some of these methods on a representative real data stream will be presented. A new Twitter data set acquired during Hurricane Florence in September 2018 is analyzed here. Instead of often-applied keyword-based filtering, only geo-located tweets from a specific area of interest were recorded. Even though a high percentage of all published tweets are not geo-tagged, this reduced set might still be a more representative subset of the complete data stream, since messages are not restricted to contain specific keywords or hashtags. In addition to first insights regarding the disaster event gained by applying analysis methods to this data set, directions for future work on that topic will be pointed out.**

I. Introduction

Twitter enables users to post tweets with their current locations shared. With an average rate of $0.85 - 3\,\%$ tweets being geo-tagged, around 7,000,000 geo-tagged tweets are posted per day [1]. A $1\,\%$ fraction of this data stream is freely available and can be accessed via the official Twitter API. Georeferenced Twitter data creates a promising opportunity for the research area of GIScience to understand geographic processes and spatial relationships inside social networks [2].

Several studies and use cases have demonstrated the great value and importance of analyzing social media streams, e.g. for rescue and helping activities in case of natural disasters and technical accidents [3]. Even though social media could be a useful information source for all phases of natural disaster management, this work focuses on disaster response.

A common approach for data stream analysis is to process the content of single tweets. Deep learning models were recently proposed for filtering crisis-related tweets [4] as well as to categorize tweets into crisis-related information classes [5], [6]. However, social media data are multi-dimensional [7]. Besides the content of social media posts, time, location and network information are additional rich and important sources of information that may help to identify, quantify, localize, map and further analyze the impact and development of natural desasters.

In this work, potential methods for analyzing Twitter data are investigated regarding spatio-temporal event impacts on population and environment. Due to the focus on natural disaster response, we are particularly interested in methods that are able to reflect important aspects of (sub-) event impacts on social media users. As an example, the spatio-temporal identification of citizens in danger of life, in temporarily inaccessible settlement areas or affected by power outages is extremely helpful for relief forces. In this regard, a workflow consisting of suitable methods for spatio-temporal Twitter data analysis is proposed. Hurricane Florence caused severe damage in North and South Carolina, USA, in September 2018. During and after this natural disaster, we acquired and analyzed $1,1$ Mio. geo-located tweets posted within the impact area. Our workflow aims on gaining a deeper understanding of how citizens use Twitter during natural disasters and what kind of hidden information can be mined. These insights may then contribute to the improvement of data analysis methods and in turn disaster risk management strategies.

The main contributions of this work are:

- State-of-the-art methods for spatio-temporal data analysis are reviewed
- A basic workflow constituted of selected methods for spatio-temporal Twitter data analysis is proposed
- The impact as well as the spatio-temporal development of hurricane Florence in September 2018 is exemplarily investigated
- Event characteristics correlated with real-world events?
- Further research directions towards the improvement of disaster risk management strategies are pointed out

The article is organized as follows. In the next section, approaches for spatio-temporal data analysis are reviewed. Utilizing suitable methods from this set, we propose a general workflow for twitter data analysis in section III. The Florence data set is introduced in section IV, followed by the application of our workflow to this data set in section V. The results are discussed in section VI. Conclusions and future research directions are pointed out in section VII.

II. Spatio-Temporal Data Analysis

A review of social media analytics in natural disaster management is provided in [7]. Approximately $1/3$ of all reviewed works focus on methods analyzing a single dimension of all potentially available features (space, time, content, and network), where $85\,\%$ focus either on content or space. In turn, $2/3$ of all reviewed papers focus on multiple dimensions. However, it is pointed out, that simultaneous analyses of multiple dimensions of social media data are rare in current literature. By analyzing more dimensions simultaneously, more information richness could be gained.

A further advanced systematic literature review on spatio-temporal analyses of twitter data is provided in [2]. The goal of reviewing 92 scientific papers is to close the gap between applied GISciences and the growing body of research works conducting Twitter data analysis, where the latter is not clearly visible and not easy to locate. According to this study, $46\,\%$ of the reviewed works deal with event detection for the application fields disaster management ($27\,\%$), disease and health management ($5\,\%$) as well as traffic management ($14\,\%$). $10\,\%$ of the reviewed papers focus on researching spatio-temporal information in Twitter not including semantic analysis. $33\,\%$ of the papers use the information layers message, geotag, and timestamp. Density-based spatial clustering techniques have been the main applied spatial methods of reviewed studies.

In [8], 31 social media visual analytic toolkits for disaster management are reviewed. For visualizing a summary of social media during crises, methods for analyzing different dimensions covering temporal, spatial, and thematic aspects, are required. $95\,\%$ of all reviewed toolkits use methods for topic/issue/trend analysis, whereas only 3 of them provide opinion or sentiment analysis. $1/3$ of the toolkits provide network analysis capabilities and $2/3$ of them provide mapping options. Furthermore, $1/3$ of the toolkits use timeline analysis and timeline visualization methods.

To summarize some of the fndings of the above mentioned studies, the following research topics were identified:

- It is necessary to put more emphasis on overcoming the computational constraints to analyze more dimensions simultaneously.
- The spatial dimension in social media data has been given particular attention, while other dimensions have not been fully exploited in data fusion.
- The fusion of social media data with census data and remote-sensing imagery is currently mainly about simple overlay and aggregation and could therefore be further investigated.
- Visual analytic tools could be developed to help determining the progress of an emergency (e.g. the rising level of flooding during a tornado or heavy rain) and to help identifying the next possible solutions (e.g. evacuation route) that could be used to alleviate the damages.

In this context, we now review suitable methods for analyzing spatio-temporal data including approaches to incorporate additional dimensions. Please note that fusing spatio-temporal data with other data sources or existing geographic knowledge as well as advanced visualizations are not in the scope of this work. Thus, several reviewed works are related to the task of event detection. Furthermore, spatio-temporal clustering, outlier detection and hotspot detection are all key research techniques in the field of spatio-temporal data mining [9]. Since clustering is often involved, common parametric or non-parametric techniques are discussed at first. Furthermore, methods for event detection as well as for spatio-temporal process analysis are reviewed.

A. Clustering Approaches

Kernel density estimation (KDE) methods estimate a density surface from a set of point-based locations by adding functions, e.g. Gaussians, centered at each data point. In [10], the spatial KDE approach is extended to incorporating temporal attributes of point data. Tackling the main problem of different units in space and time, the authors propose different extensions of KDE (space-time KDE, STKDE). Experimental results show that STKDE methods can identify hotclusters and coldclusters better than KDE. An exemple visualization of STKDE results are shown in Fig. 1.

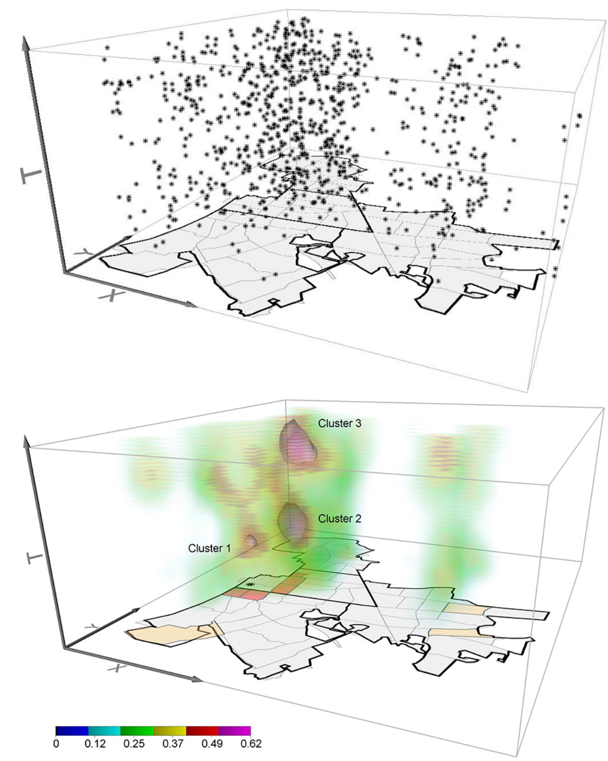

Fig. 1: Robbery Crime Locations (Top) and Corresponding Spatio-Temporal Clusters (Bottom) in Baton Rouge, Louisiana in 2010 [1].

[1]http://faculty.cas.usf.edu/yhu/

A further well-known approach for hotspot analysis is the *Getis-Ord G** method [11] that detects local attribute clusters in data. This approach for local spatial autocorrelation is applied for the purpose of disaster footprint estimation and damage assessment based on social media data in [12]. Example results for time-aggregated hotspot analysis results are shown in Fig. 2.

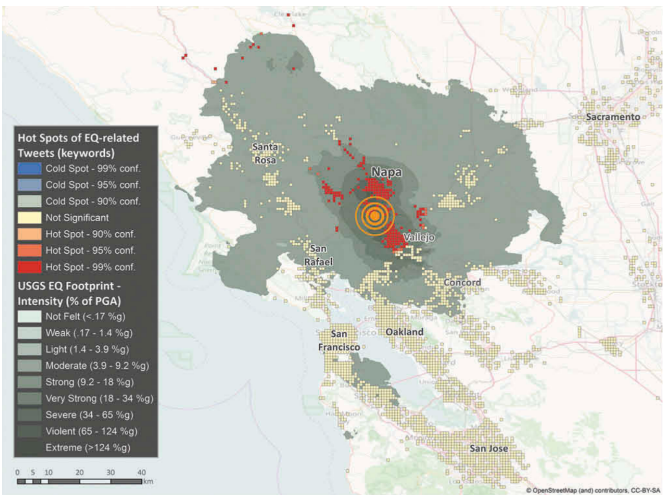

Fig. 2: Keyword-based earthquake-related Tweet hot spots [12]. ©OpenStreetMap (and) contributors; available under CC-BY-SA Licence

Space time scan statistics (STSS), implemented in the freely available software SaTScan[2], is a further approach to analyze data points (incidences) within a space-time cube. A cylindrical window, of varying radius (space) and height (time) is moved across all possible space-time locations. Based on the number of observed incidents compared to the number of expected incidents, the clusters of interest are identified from the set of all candidate clusters. Each clusters significance is then tested, giving each a p-value, describing the likelihood that it occurred by chance (see [13] for further details). An exemplary STSS result is shown in Fig. 3

DBSCAN [14] is a further spatial density-based clustering algorithm for applications with noise. The number of clusters is identified based on the quantity of highly density connected components. Parameters are the radius and the minimum number of neighbors. From these parameters, clusters with different formats and the same density, are found. ST-DBSCAN [15] is an extention of DBSCAN discovering clusters according to non-spatial, spatial and temporal features. In [1], this approach is utilized for the spatio-temporal clustering of tweets recorded during gunshot events. An example result is shown in Fig. 4.

Gaussian mixture modeling is a further approach for clustering spatial as well as spatio-temporal data. The advantage of such a model-based approach is that no decisions have to be made about the scaling of variables. Mixing different units is allowed as well. Furthermore, since mixtures of different

[2]https://www.satscan.org/

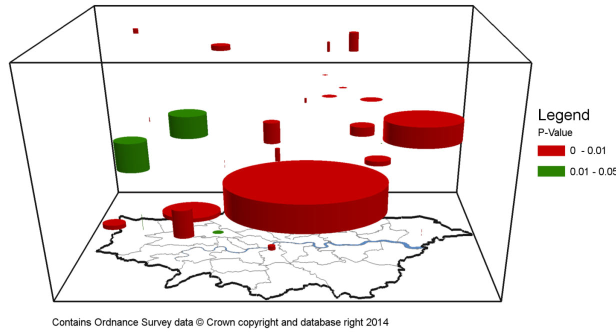

Fig. 3: Analyzing Twitter activities in case of a helicopter crash: Significant hourly clusters within London between 16th January and 17th January 2013 [13]. Contains Ordnance Survey data ©Crown copyright and database right 2014.

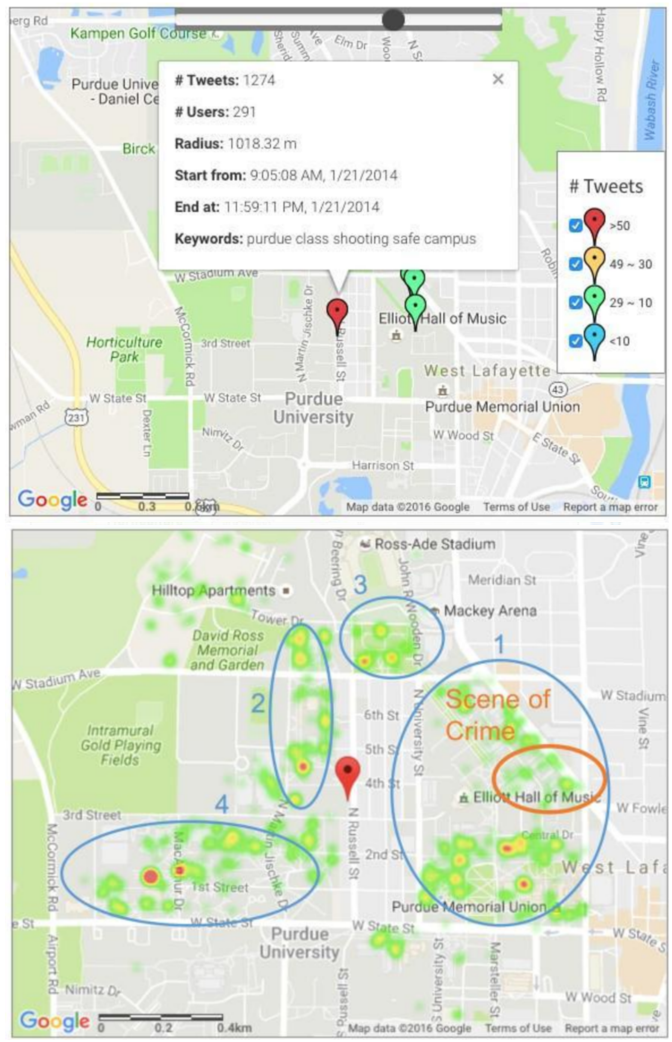

Fig. 4: Spatial-temporal clustering results for the tweets in West Lafayette on 21 January 2014 when a gunshot occurred: (top) cluster centers and the most significant cluster; (bottom) distribution of the tweets in the most significant cluster [1].

distributions are allowed, the joint analysis of continuous and categorial data is enabled. Thorough surveys of state-of-the-art mixed data clustering algorithms are provided in [16], [17].

In [18], different methods for unsupervised topic modeling for short texts using distributed representations of words, namely Gaussian mixture models (GMM), probabilistic latent semantic analysis (pLSA) and latend Dirichlet allocation (LDA), are compared. To overcome the problem of rather sparse word co-occurrence statistics, a vector space model for representing words (embedding) is used here. The authors conjecture that the GMM can learn the latent topics by clustering over the distributed representations trained with a semantic similarity objective. The experimental results indicate that this approach can reliably learn latent topics and can be used to categorize short messages with high fidelity in comparison with LDA and biterm topic models. However, experiments conducted in [19] revealed that biterm models provided better results in modeling topics of tweets. One reason for the observed poor performance of GMMs in their study might be the assumed feature independency (diagonal covariance matrices).

With special emphasis on numerical and categorial data, a fast density clustering algorithm (FDCA) is proposed in [20]. Cluster centers automatically determined by a so-called *center set algorithm* (CSA) have to be scanned once. A novel data similarity metric is designed for clustering data including numerical and categorical attributes. CSA is designed to choose cluster centers from data object automatically which overcome the cluster centers setting difficulty in most clustering algorithms.

B. Event Detection

A recent survey related to event detection in Twitter data is provided in [21]. Here, different supervised and unsupervised methods for the detection of physically occuring events as well as for emerging/popular topic detection are discussed. Furthermore, a survey on real-time real-world event detection from the Twitter data stream is conducted in [22]. Methods are grouped into term-interestingness-based, topic-modelling-based, incremental-clustering-based and miscellaneous approaches. The authors conclude, that further work is required to propose effective measures to filter out spam and trivial events. In addition, it is imperative that proposed event detection systems are evaluated on a publicly available text corpus of data, with a rich collection of a variety of events, to facilitate a fair comparison of different systems.

In [13], STSS is applied for event detection using Twitter data. STSS looks for clusters within the dataset across both space and time, regardless of tweet content. It is expected that clusters of tweets will emerge during spatio-temporally relevant events, as people will tweet more than expected in order to describe the event and spread information. A spatio-temporally significant cluster is found relating to a London helicopter crash. Although the cluster only remains significant for a relatively short time, it is rich in information, such as important key words and photographs. The method also detects other special events such as football matches, as well as train and flight delays from Twitter data. These findings demonstrate that STSS is an effective approach to analysing Twitter data for event detection.

With a special emphasis on real-world events, an approach for geo-spatial event detection in the Twitter Stream is proposed in [23]. Target events are often expected on a rather small-scale, meaning that they happen at a specific place in a given time period, and are often covered by only few tweets. For the processing of more than three million tweets acquired during one day, a pre-selection of tweets based on their geographical and temporal proximity is conducted. Clusters (event candidates) are constituted based on simple distance measures in space and time. 41 features that address various aspects of the event candidates are computed for each tweet. These are used to rank the tweets and to make a binary decision as to whether a tweets cluster constitutes a real-world event or not. Processing is done in real-time, and clusters classified as events are shown to a user in a GUI.

C. Spatio-Temporal Process Analysis

In this section, methods and frameworks to analyze the spatio-temporal development of processes are reviewed.

The detection of small-scale spatio-temporal events from geo-tagged tweets is focused in [1]. ST-DBSCAN is utilized to spatially-temporally cluster tweets. After summarization of the word frequencies for each cluster, potential topics are modeled by LDA. ST-DBSCAN has shown to be a good method for the detection of several different small-scale event types, e.g. gunshot events (fig. 4). However, a careful selection of the four cluster parameters is required. Furthermore, the authors introduced two further parameters to further enhance the clustering results. Based on various examples it could be demonstrated that Twitter can be utilized to discover real-world events as well as to understand their intrinsic behaviors. For further reserach, the authors recommend to analyze UTF-8 characters, such as emojis, or languages other than English. For events that last for several days, extra effort is required for combining data. Furthermore, a better understanding of textual information can facilitate the interpretation of the events.

An approach based on machine learning and geovisualization to identify events in cities and trace the development of these events in real-time based on Twitter data is presented in [24]. Twitter data is stored in MongoDB documents, enabling spatial and textual queries. After tweet pre-processing, including stop-word removal and tokenization, the resulting word sets are grouped into one-our intervals. In order to find tokens related to events, bursty word detection techniques are applied. Bursty words are spikes in the frequency of tweets along the time spectrum. DBSCAN is utilized for spatial clustering and Random Forests were trained to classify tokens as event-related or unrelated. One interesting feature of this approach is that it does not assume any prior knowledge about events. On the other hand, only major local events can be revealed, due to the low proportion of geo-tagged tweets. Instead of using single tokens, n-grams might be used to better represent longer expressions.

A statistical approach for studying the spatio-temporal distribution of geolocated tweets in urban environments is proposed in [25]. Instead of detecting and characterizing unique events, repeated patterns are focused here. A negative binomial regression analysis for the time series of counts of tweets is the first step. Then, a functional principal component analysis (FPCA) of second-order summary statistics of the hourly spatial point patterns formed by the locations of the tweets is applied. Finally, groups of hours with a similar spatial arrangement of places where humans develop their activities through hierarchical clustering over the principal scores are identified. Social media events are found to show strong temporal trends such as seasonal variation due to the hour of the day and the day of the week. Furthermore, spatio-temporal patterns of clustering, i.e., groups of hours of the day that present a similar spatial distribution of human activities, are identified. This approach may help to characterize the usual Twitter usage behavior as a starting point for anomaly or event detection.

In [26], a method for discovering the spatio-temporal process of a typhoon using Sina Weibo micro blog data is presented. SVMs are utilized for classifying text messages into four classes: *warning information*, *disaster information*, *irrelevant information* and *rescue information*. Considering the spatial heterogeneity of Weibo users, a weighted model based on user activity at the check-in points is proposed. Further spatio-temporal analyses are based on searching keywords. The authors point out, that more research is required to determine the most appropriate scale for spatial analyses. Furthermore, deep learning should be involved, e.g. to detect lexical similarities.

In [12], machine-learning topic models (LDA) and spatio-temporal analysis (local spatial autocorrelation) of social media data are combined for hotspot detection, disaster footprint estimation and damage assessment. LDA is applied to a twitter data stream in a cascading fashion in order to extract earthquake- and damage-related tweets. After statistical topic validation, hotspot analysis is conducted based on local spatial autocorrelation (*Getis-Ord G**). In order to mitigate effects of strongly accumulated tweet occurrences in urban areas (making it impossible to draw conclusions on the location of the earthquake and on potential damage), earthquake-related tweets are normalized (1) over the population by using the LandScan population layer[3] at a resolution of 800 m and (2) the overall number of tweets per cell. Topics for future work are optimization of the preprocessing routines, automated interpretation of the generated semantic topics, development of a standardized method for determining the best parameters for the semantic analysis, dealing with the lacking representativeness (population-wise and spatially) of social media posts, the creation of an automated validation procedure, and the development of a real-time monitoring system.

[3]http://web.ornl.gov/sci/landscan/

In [27], the spatio-temporal variability in social media response is examined and an approach to leverage geotagged tweets to assess the evacuation responses of residents is developed. The approach involves (keyword- and location-based) retrieval of tweets, creation and filtering of different datasets, and statistical and spatial processing to extract, plot and map the results. Spatial and temporal analyses, which are done separately in this study, are mainly based on counting related tweets (per time and/or per county). Evacuation-related analyses, i.e., how many Twitter users were evacuated, their evacuation destination, and return date, involves the identification and tracking of active local users during different event periods.

In contrast of directly using Twitter data, a framework for discovering evolving *domain related* spatio-temporal patterns is proposed in [9]. Given a target domain, a dynamic query expansion is employed to extract related tweets which are then used to form spatio-temporal Twitter events. The new spatial clustering approach proposed here is based on the use of multi-level constrained Delaunay triangulation to capture the spatial distribution patterns of Twitter events. An additional spatio-temporal clustering process is then performed to reveal spatio-temporal clusters and outliers that are evolving into spatial distribution patterns. Future work will focus on the analysis of quality, incompletion and uncertainty Twitter data.

A method for spatio-temporal anomaly detection through visual analysis of geolocated Twitter messages is proposed in [28]. The approach enables the interactive analysis of location-based microblog messages in realtime by means of scalable aggregation and geolocated text visualization. For this purpose, a novel cluster analysis approach is used to distinguish between local event reports and global media reaction to detect spatio-temporal anomalies automatically. The main tool for data analysis and visualization is the so-called ScatterBlogs system. Its purpose is to enable analysts to work on quantitative as well as qualitative findings by not only automatically identifying anomalies, but also summarizing and labeling event candidates as well as providing interaction mechanisms to examine them. To generate an overview of potential anomalies from Twitter data streams, three activities are proposed: (1) Extraction of message terms and transformation to term artifacts, (2) creation of spatio-temporal clusters by quantization of term artifacts, (3) selection of clusters as anomaly candidates by a decision strategy. The identified anomalies are then represented as term map overlays for exploration.

In [29], a spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation is proposed. The framework has four major features: (1) a spatio-temporal kernel density estimation (STKDE) method is applied to include the temporal component in predictive hotspot mapping, (2) a data-driven optimization technique (likelihood cross-validation) is used to select the most appropriate bandwidths, (3) a statistical significance test is designed to filter out false positives in the density estimates, and (4) a new metric is proposed to evaluate predictive hotspots at multiple areal scales. The applied STKDE method treat space and time as independent components, while neglecting the spatio-temporal interactions in the process. One remedy is to examine if there are any spatio-temporal dependency patterns in the data before applying STKDE. A more meaningful way is to design a STKDE that considers such dependency patterns in the model.

D. Discussion

The goal of this work is to identify and test potential methods for analyzing Twitter data regarding spatio-temporal impacts of natural disaster events. Exemplarily, we focus on a Twitter data set recorded during hurricane Florence, in September 2018. According to [9], one might distinguish between (1) spatio-temporal distribution pattern detection from initial Twitter data; and (2) spatio-temporal distribution pattern detection from domain related Twitter events. Proposing a workflow for the latter case, the authors use a dynamic query expansion to extract domain related tweets from Twitter [30]. Instead, the use of recently proposed deep learning methods for extracting either disaster-related tweets [4], [31], or more-specifically, event-related Tweets [32] might be a promising alternative. In the first case, one might identify and characterize disaster events in general, while in the latter case the detection of new sub-events during a known event is desired.

The application of state-of-the-art methods to classify crisis-related tweets into more finely-grained information classes [5], [6], [33], [34] provides additional valuable information that should be combined with spatio-temporal analyses. A good starting point for several analyses might be to utilize simple statistical analyses[4,5] [11], as well as state-of-the-art methods, like STKDE[6,7] [10] and ST-DBSCAN[8] [14].

As pointed out in [7], the analysis of multiple dimensions should be preferred over analyzing single dimensions. In particular, the direct incorporation of semantic analyses might be quite interesting. One approach to tackle this could be to incorporate classification labels obtained for single tweets, for example by using tweet filtering or classification methods. Instead of class labels, one might also directly use vector representations (embeddings) to represent high-frequency terms or tweets as low-dimensional vectors [18], [19] and utilize methods, like EM [17] or FCDA [20] to cluster the data. Using EM, the dependency of dimensions should not be neglected. Pitfalls of EM are the dependency on initialization as well as the usually unknown number of mixture components. One solution for these two problems is provided in [35]. While the relevant feature dimensions are additionally estimated during EM, the aforementioned independence assumption is introduced in order to make the computations tractable. Both approaches, EM and FCDA are well suited for clustering numerical and categorial features simultaneously and are therefore good candidates for further research. Whereas implementations of EM [36]–[39] are freely available [9,10,11].

III. Proposed Workflow

Based on the above-discussed aspects, an initial workflow for spatio-temporal Twitter data analysis in context of natural disasters is proposed in this section. At the current state of or research, the analysis of Twitter data recorded in the past, for example from the last few days, is focused. Hence, the two main modules *data acquisition* and *data analysis* are conducted sequentially (see figure 5).

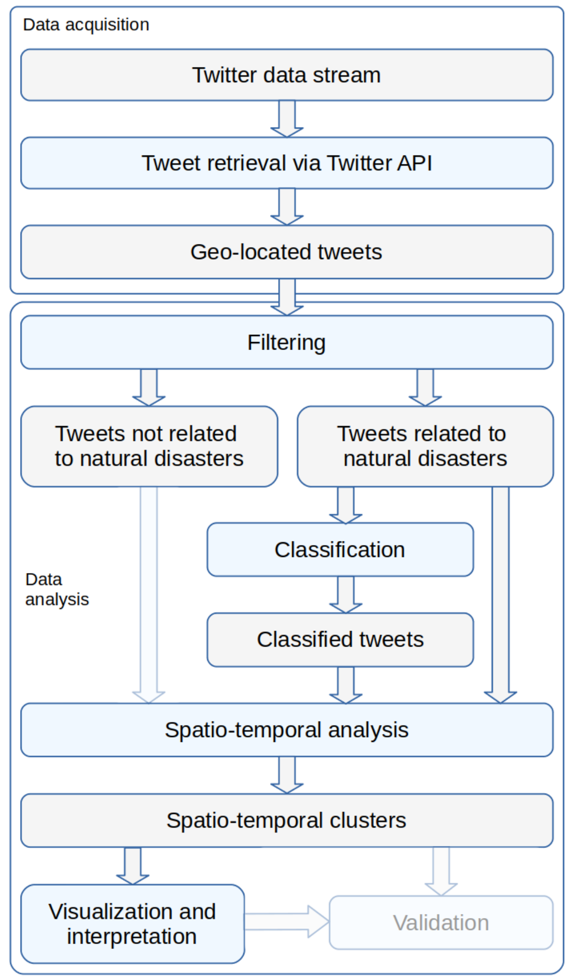

Fig. 5: Workflow for spatio-temporal Twitter data analysis. The desaturated components are not covered in this work.

A 1 % fraction from the full Twitter data stream can freely be accessed using the official Twitter API. A restriction to tweets with a geo-location is necessary for spatio-temporal analyses related to real events. Even though this further reduces the amount of tweets, the contributions reviewed above have shown that this portion of data provides rich

[4]https://github.com/maovieira/localGi
[5]https://github.com/danioxoli/HotSpotAnalysis_Plugin
[6]https://github.com/karpfen/stkde
[7]https://github.com/alexandster/densitySpaceTime
[8]https://github.com/eubr-bigsea/py-st-dbscan

[9]www.maths.uq.edu.au/gjm/emmix/emmix
[10]https://www.stat.washington.edu/raftery/Research/Mclust/mclust.html
[11]https://swmath.org/software/3250

and significant information that is useful in many application domains. Since the amount of messages related to disaster events is usually small, a filtering step is conducted in order to identify the tweets related to disaster events. Similar to [9], subsequent analysis steps can be seen as a domain-related approach.

For identifying tweets related to any natural or man-made disaster, the approach proposed and thoroughly investigated in [31] is utilized in this work. With this Convolutional Neural Network (CNN) model, F1-scores of 0.83 ± 0.13 can be expected in case of unseen events. Alternatively, one might be interested in identifying tweets related to a specific event that just started to emerge. In this case, recently proposed few-shot models [32] may be a reasonable alternative. The advantage of these pre-trained CNN models is, that they can be efficiently applied to new events by specifying only a few (for example 50, but also 10 might be sufficient), representative tweets. Since this type of few-shot model tackles one-class problems, the much more complex distribution of the class of tweets not related to the specific events has not to be taken into account. We then apply another CNN proposed in [45], in order to classify the related tweets into information classes, like *affected individuals* or *infrastructure and utilities*.

In the next step, the disaster-related tweets, comprising text, timestamp, geo-location, as well as the assigned information class, are then spatially-temporally analyzed. In order to gain first insights into our data set, the aforementioned and well-known approaches Getis-Ord G*-statistics, STKDE, and ST-DBSCAN are exemplarily utilized in this study.

Following the recommendation of jointly analyzing multiple dimensions in [7], alternative approaches, where either the class labels or text message vector representations (embeddings) are used as additional features, seem to be promising alternatives, but are not taken into account here. Furthermore, tweets classified as unrelated might additionally be used, for example to derive an approximate mean distribution of tweets over time and space. This knowledge can help to gain further insights regarding the typical occurence of Twitter messages within a specific region over time and in turn may contribute to the detection of anomalies. This interesting aspect might be part of our subsequent research.

The resulting spatio-temporal clusters as well as their visualization need then to be interpreted. A validation of the obtained spatio-temporal clusters can easily be done when a dataset with ground-truth is used. In case of new data, one might find correlations between detected spatio-temporal clusters and known (sub-) events, e.g. reported in the news or Wikipedia. The quantitative validation of results will also be part of future work and is therefore marked with an increased transparency. In the following, the applied methods for spatio-temporal analyses are described in detail.

A. Getis-Ord G-statistics

This statistic measures the degree of association that results from the concentration of weighted points (or area represented by a weighted point) and all other weighted points included within a radius of distance d from the original weighted point [11]. Hence, the G-statistic is a tool for hot- and cold-spot analysis, in which high and low values are clustered and the concentration of these values in a specific area is measured. The general G_i-statistic of a specific point i is calculated by

$$G_i = \frac{\sum_{j=1}^{n} w_{i,j} x_j}{\sum_{j=1}^{n} x_j}, \forall j \neq i \tag{1}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between entity i and j, and n is the number of entities (e.g. tweets).

The G_i^* statistic, in which the case $i = j$ is also allowed, is computed as follows

$$G_i* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}}, \tag{2}$$

with

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{3}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{X}^2}. \tag{4}$$

An entity with a high value might be interesting, but may not be a statistically significant hot spot in terms of the G_i^* statistic. To be a statistically significant hot spot, an entity will have a high value and be surrounded by other features with high values as well. The local sum for an entity and its neighbors is compared proportionally to the sum of all entities. When the local sum is much different than the expected local sum, and that difference is too large to be the result of random chance, a statistically significant G_i^* value (Z score) results.

B. STKDE

Given a spatial distribution of geographic events in a region, KDE methods compute a density surface that describes the intensity of the geographic events' distribution. KDE can be understood as a simple adding of basic density functions, for example Gaussians, located at each event location. Hence, in contrast to interpolation methods, in which attribute values at various locations are determined, intensities of spatial distributions of point data are estimated in the following way:

$$\hat{f}(x, y) = \frac{1}{n h_s^2} \sum_{i=1}^{n} k_s \left(\frac{x - x_i}{h_s}, \frac{y - y_i}{h_s} \right), \tag{5}$$

where n is the number of data points, x_i and y_i define the location of point i, h_s is the bandwidth of the kernel density function, and k represents the kernel density function.

Space-time kernel density estimation (STKDE) is an extension of the traditional KDE, widely used for identifying spatiotemporal patterns of underlying datasets through visualizing the resulting densities in the space-time cube. The output of STKDE can be expressed as a 3D raster volume, where for each voxel a density estimate based on the surrounding point data is assigned. The corresponding density values can be computed with the following equation:

$$\begin{aligned}\hat{f}\left(x,y,t\right) =&\frac{1}{nh_s^2 h_t}\sum_{i=1}^{n} I\left(d_i < h_s, t_i < h_t\right)\\ & k_s\left(\frac{x-x_i}{h_s},\frac{y-y_i}{h_s}\right) k_t\left(\frac{t-t_i}{h_t}\right),\end{aligned} \tag{6}$$

where the density $\hat{f}$ is computed based on all points (x_i, y_i, t_i) for which the spatial distance d_i and the temoral distance t_i to the current point is lower than the thresholds d_i and t_i, respectively. Hence, the indicator function I returns 0, if one of these thresholds is exceeded and otherwise 1. For the kernel functions k_s and k_t, the Epanechnikov kernel [40] is applied here, where each data point is weighted according to its distance in time (t_i) and space (d_i) to the current voxel (the closer the data point, the higher the weight).

C. ST-DBSCAN

ST-DBSCAN [15] is a clustering approach based on DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [14]. In DBSCAN, the density associated with a point is obtained by counting the number of points in a region of specified radius around the point. Points with a density above a specified threshold are constituted to clusters. Compared to so-called parametric approaches, no assumption about the cluster shapes is introduced. Hence, DBSCAN has the ability in discovering clusters with arbitrary shape such as linear, concave, oval, etc. Furthermore the number of clusters is estimated.

ST-DBSCAN (figure 6) can cluster spatialtemporal data according to non-spatial, spatial and temporal attributes. Besides a distance measure (Euclidean distance) between data points in DBSCAN, a second similarity measure is introduced, in order to describe non-spatial similarity between attributes, like temerature or time. To speed up computations, similarity measures are only comuted for temporal neighbors. If a point does not contribute to any of the clusters (i.e., the number of neighbors is lower than a threshold), it is considered to be a noise point.

IV. Data Set: Hurricane Florence

We acquired $\sim 600,000$ tweets during hurricane Florence from September 12th to 19th 2018, using the Twitter API. For location-based filtering, the area of interest (AOI) shown in figure 7 was defined. Forecast services estimated that this would be the mainly affected area for which heavy rainfall was expected as well (figure 8).

```
Algorithm ST_DBSCAN (D, Eps1, Eps2, MinPts, Δε)
  // Inputs:
     // D={o1, o2, ..., on}  Set of objects
     // Eps1 : Maximum geographical coordinate (spatial) distance value.
     // Eps2 : Maximum non-spatial distance value.
     // MinPts : Minimum number of points within Eps1 and Eps2 distance.
     // Δε : Threshold value to be included in a cluster.
  // Output:
     // C={C1, C2, … Ck} Set of clusters

 Cluster_Label = 0

 For i=1 to n                                   //(i)
   If oi is not in a cluster Then               //(ii)
     X=Retrieve_Neighbors(oi , Eps1, Eps2)      //(iii)

     If |X| < MinPts Then
          Mark oi as noise                      //(iv)
     Else                        //construct a new cluster (v)
        Cluster_Label = Cluster_Label + 1

        For j=1 to |X|
          Mark all objects in X with current Cluster_Label
        End For

        Push(all objects in X)                  //(vi)

        While not IsEpmty()
           CurrentObj = Pop()
           Y= Retrieve_Neighbors(CurrentObj, Eps1, Eps2)

           If |Y| >= MinPts Then
              ForAll objects o in Y             //(vii)
                 If (o is not marked as noise or it is not in a cluster) and
                    |Cluster_Avg() - o.Value| <= Δε Then
                    Mark o with current Cluster_Label
                    Push(o)
                 End If
              End For
           End If
        End While
     End If
   End If
 End For
End Algorithm
```

Fig. 6: ST-DBSCAN algorithm [15].

Fig. 7: Defined AOI [41]

Since only a fraction of all tweets is known to be georeferenced, this approach clearly introduces bias. On the other hand, filtering by keywords might be problematic as well, since unrelated messages containing the keywords will be retrieved, while important messages not containing any of the keywords will be discarded. Besides the fact that the geo-location is required for our analyses, this approach provides tweets from directly affected individuals, whereas tweets from users who are not directly involved, but contribute to discussions, are discarded.

V. Results

In this section, the results obtained with the proposed workflow applied to the hurricane Florence data is presented.

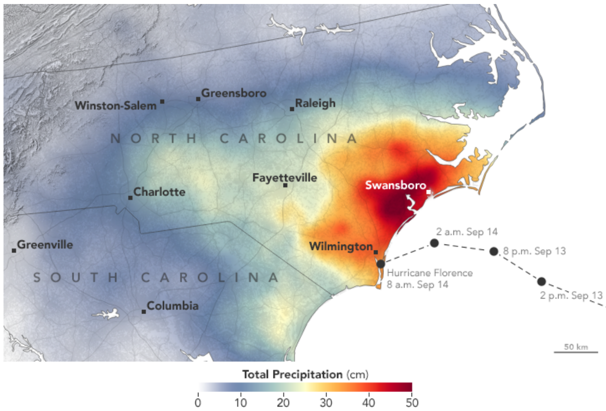

Fig. 8: Rainfall and hurricane trajectory (Sep. 13-14) [42]

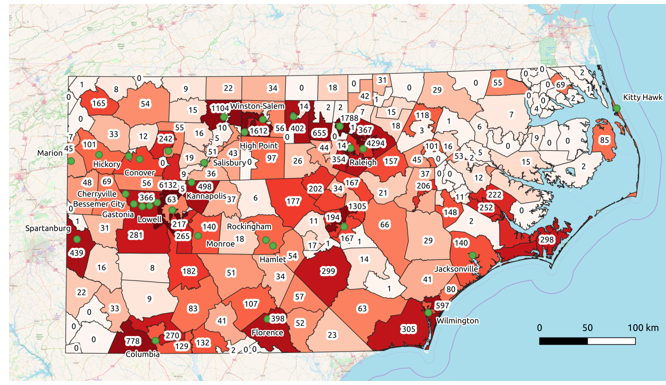

Fig. 9: Number of geo-located tweets per county/congressional destrict. Dark red areas correspond to a high number of tweets (non-linear scaling).

A. Filtering of Unrelated Tweets

According to figure 5, the first processing step after retrieving the geo-located tweets is filtering, i.e. distinguishing between crisis-related or unrelated messages. The total number of geo-located Tweets recorded during hurricane Florence is around 600,000. Our CNN-based filtering approach identified approximatedly 88,000 crisis-related tweets from this initial set (see figure 10). Experimental results conducted with labeled data as well as with the Florence data set in [31] revealed, that the applied CNN model tends to provide high precision and lower recall, where the false positives tend to have low CNN model likelihoods.

Based on this these properties, around 32,000 tweets with a likelihood below 0.7 are therefore removed from the set of relevant tweets. Furthermore, around 25,000 tweets with a county-level geo-location have to be excluded from spatio-temporal analyses, since their coordinate is set to the centroid of the corresponding US county and is therefore not representative. Note that even this significant fraction of messages is removed for the preceding spatio-temporal analyses, but shpuld in general be taken into account. The spatial distribution of the resulting number of around 30,700 tweets is shown in 9.

B. Spatio-Temporal Analyses

1) Getis-Ord G-statistics: We determine spatial hot- and cold-spots in the filtered, i.e., crisis-related, Florence data by using the Getis-Ord G*-statistics. As pointed out in [12], useful spatial information cannot be derived without taking into account the population density. We therefore normalized the number of tweets (figure 9) by the number of population per square kilometer for each congressional destrict [43] and determined the G*-statistic with a spatial bandwidth of $100\,km$ (see figure 11). As expected, the number of posted tweets and therefore also the number of disaster-related tweets, is high in areas with a high population density. On the other hand, the density is quite low in the southern and western area. The choice of a suitable bandwidth for the 30,000 tweets is therefore not an easy task. The data sparsity might be the reason that analyzing time slices, e.g. all tweets per day, did not work well with our data.

Whereas the western region around Charlotte, as well as the southern region around Wilmington had relatively high tweet densities, they were identified as potential cold-spot and not significant hot-spot, respectively. The reason for this is of course the high population density, especially around Charlotte. The identified northern hot-spot region covers the cities Greensboro, Burlington, Durham, Cary, Raleigh, Fayetville and Pinehurst. An interesting observation related to the impact of a high bandwidth is, that even if the normalized number of posted tweets is zero, the G*-score can be high, if the surrounding G*-scores are high as well. An example for this is a horizontally centered, northern region near to Burlington, which is surrounded by areas with tweet counts of 34 (west), 14 and 402 (south) and 18 (east) in figure 9.

To summarize, the G*-statistics can be used to identify high- and low-activity areas, where Twitter users post messages related to crisis events. Potential hot- and cold-spots are determined by proportionally comparing the local sum for a tweet and its neighboring crisis-related tweets to the sum of all crisis-related tweets. An interesting outcome of the analysis is, that the area around Charlotte - the city with the highest population in North Carolina - is identified as a cold-spot, even though the number of identified crisis-related tweets in this region is similar as in the identified hot-spot region. Whereas this clearly helps to identify hot-spots, i.e., high activity regions related to a specific topic, the regions with high population density might offer a similar amount of information that should be considered for relief activities. Furthermore, the investigated AOI, the distribution and sparsity of data as well as the chosen spatial bandwidth have a huge impact on the results.

2) STKDE: In contrast to the G*-statistics, STKDE might be directly applied to data points, i.e., crisis-related tweets, in the (x, y, t) domain. Even though information regarding the population are not taken into account with this approach, the resulting density estimates are meaningful, in case a specific

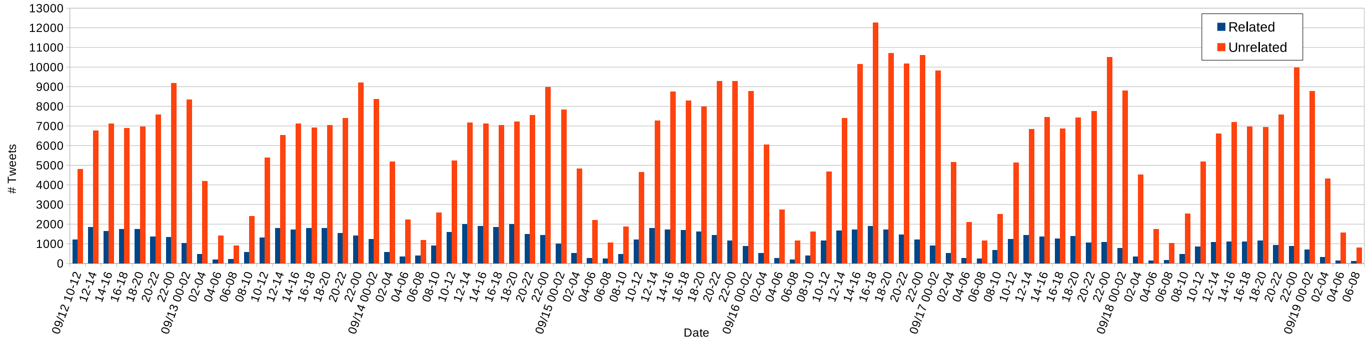

Fig. 10: Filtering results during hurricane Florence, September 12-19, 2018

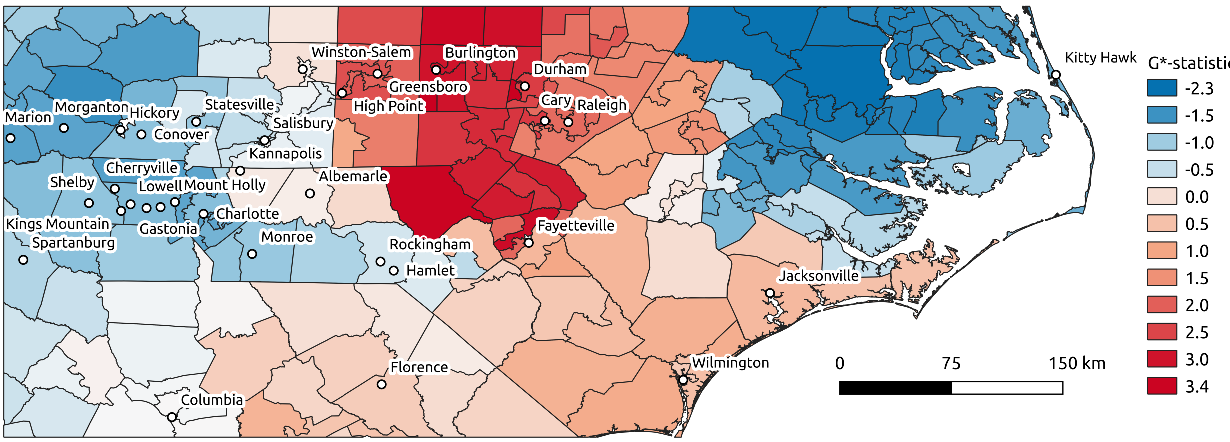

Fig. 11: Getis-Ord G*-statistics on county/congressional district level obtained with a bandwidth of 100 km. Hot- and cold-spots correspond to blue and red colors, respectively.

subset (crisis-related) of all tweets is analyzed. Since the approach descibed in section III-B is applied here, density estimates are computed for each point in a regular grid with a chosen spatial resolution of 5 km and a temporal resolution of 0.1 days. Optimal values of $h_s = 7.5\,km$ and $h_t = 0.3\,d$ for the spatial and temporal bandwidth were empirically identified. A 2D-view of the resulting dense disaster-related tweet regions by STKDE (over the whole period of our data set) is shown in figure 12. The points are color-coded and their size is increased according to the cluster density (normalized to $[0, \ldots, 1]$). The areas around Charlotte and Raleigh are identified as dense crisis-related tweet activity areas, where less dense areas are identified for example in Durham, Fayetteville, Columbia and Wilmington.

It is worth noting that the 2D-view in figure 12 does not represent the result of a 2D-analysis in which all disaster-related tweets from different points in time are projected into the 2D-plane. A 3D-view on the very same STKDE result in the spatio-temporal domain is depicted in figure 13. In this plot, spatio-temporal clusters corresponding to specific regions as well as temporal periods (mainly per day) can be observed. According to this, a clearly visible concentration of crisis-related tweets can be observed, for example in the region around Charlotte as well as Raleigh and Durham. The hurricane's landfall occured in the morning of September 14. During the two days before the landfall, approximately constant daily tweet activity patterns can be observed. Around the landfall time, a significant increase of crisis-related tweets can be observed. From this point on, the amount of tweets statically decreases towards the end of the recording period. A second peak in tweet activity can be observed on September 16, in Charlotte. A further increased tweet activity is visible in the Durham area at September 17.

The presented STKDE results demonstrate that this is a quite useful approach for visualizing and analyzing spatio-temporal distributions of crisis-related twitter activities. Similar to the G*-statistics, the choice of parameters turns out to be difficult. To solve the problem that spatial patterns of social media activities is often a reflection of the population distribution, a dual KDE approach, proposed in [44], might be applied instead.

The ST-KDE results represent an estimate of the probability density function of crisis-related tweets, which is helpful for visual analyses. However, the tweet content shall further be analyzed in order to gain more detailed insights about the discussed topics. Spatio-temporal term usage [28], unsupervised topic modelling [12], [18] or the classification of crisis-related tweets into semantic classes [34] might be suitable directions

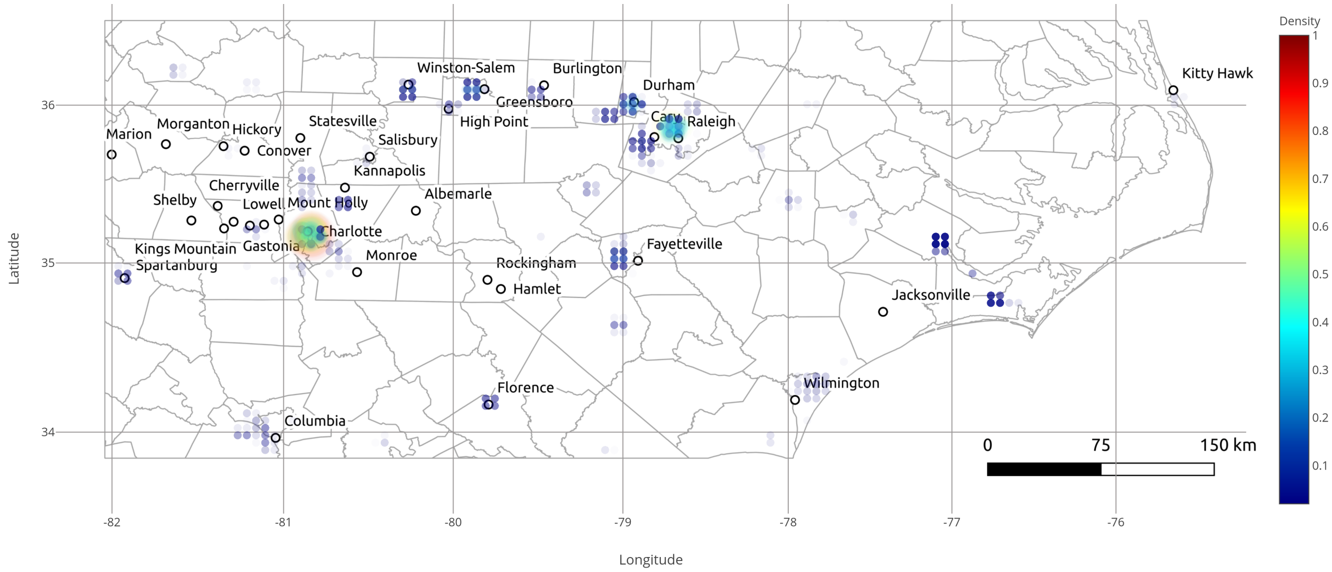

Fig. 12: 2D plot of STKDE result obtained with spatial bandwidth: $7.5\,km$, temporal bandwidth: $0.3\,d$, spatial resolution: $5\,km$, and temporal resolution: $0.1\,d$.

for this.

3) ST-DBSCAN: For ST-DBSCAN, we used similar parameters compared to the other two approaches, i.e., a spatial threshold of $10\,km$ and a temporal threshold of $1\,h$. Furthermore, the minimum number of points was chosen to be 5. If the spatio-temporal number of neighbors within a spatial radius (spatial threshold) of a tweet is below this value, the corresponding tweet is classified as noise and therefore does not contribute to a cluster.

Color-coded scatter plots of the identified clusters are shown in figure 14 and 15. In order to avoid color similarities for neighboring clusters with similar IDs, the color for each cluster ID is randomly chosen. A total number of 435 spatio-temporal clusters are identified with the above-mentioned method parameters. Furthermore, 4,785 tweets do not belong to any of these clusters and are therefore classified as noise. In contrast to the STKDE density plots (figures 12 and 13), ST-DBSCAN determines the cluster membership for each point without interpolating values for each discrete position of a regular space-time volume.

Whereas the STKDE densities reflect the amount of crisis-related tweets for specific positions and times, ST-DBSCAN produces spatio-temporal clusters that might not necessarily correspond to a rather broad region and a typical activity interval (typically one day).

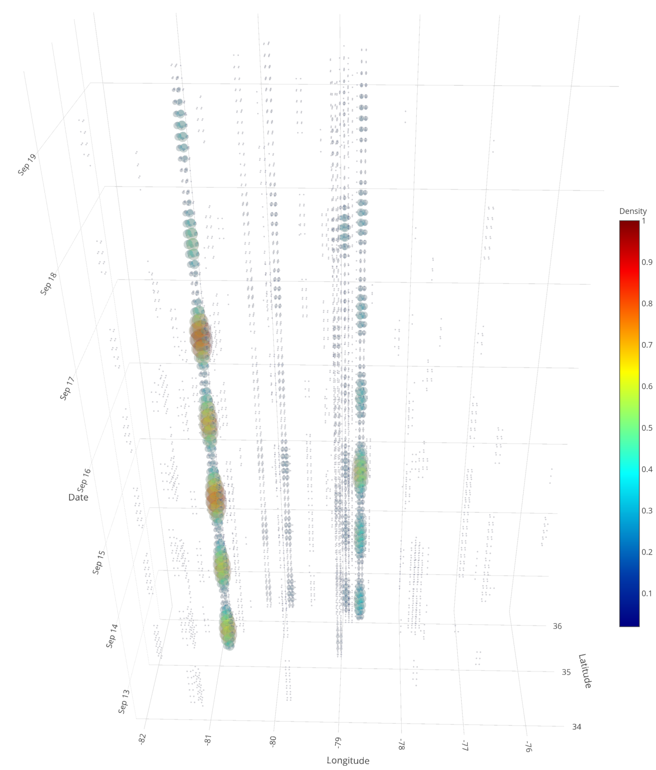

Fig. 13: 3D view on STKDE result obtained with spatial bandwidth: $7.5\,km$, temporal bandwidth: $0.3\,d$, spatial resolution: $5\,km$, and temporal resolution: $0.1\,d$.

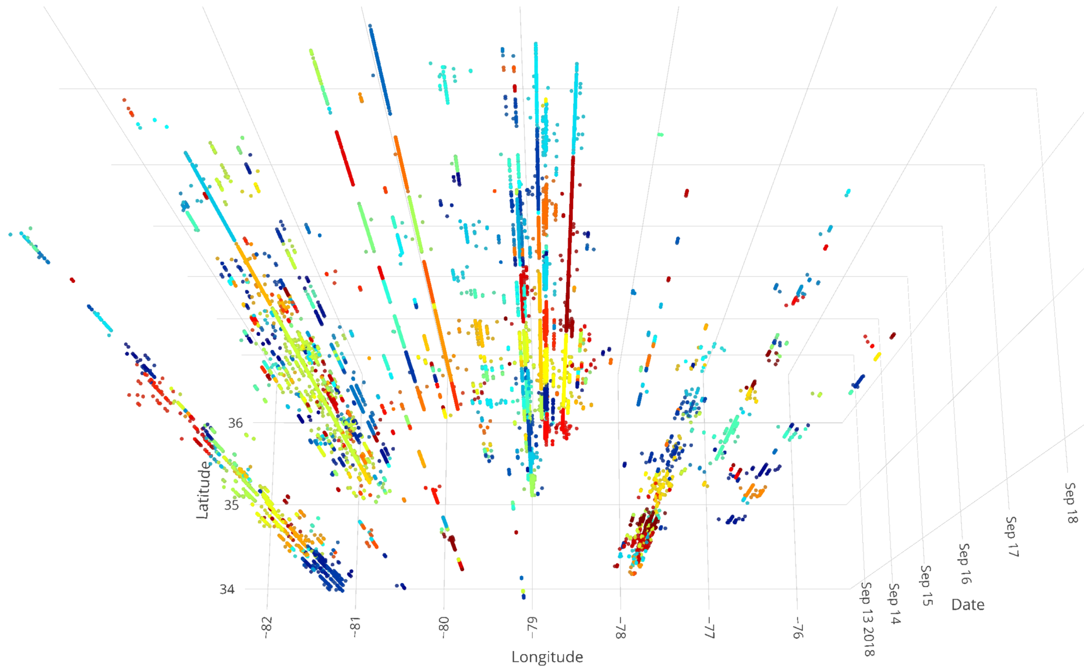

Fig. 14: ST-DBSCAN result obtained with spatial threshold=10 km, temporal threshold=1 h, minimum number of points=5. Each cluster is randomly colored.

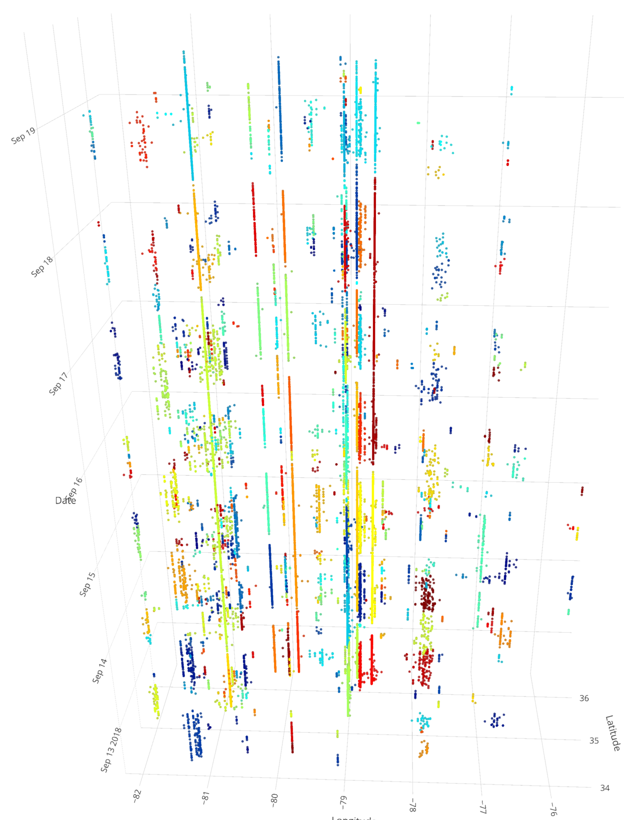

Fig. 15: ST-DBSCAN result obtained with spatial threshold=10 km, temporal threshold=1 h, minimum number of points=5. Each cluster is randomly colored.

Besides those clusters representing the complete amount of tweets per a single day in a specific region, both, multiple clusters per day and region, as well as single clusters covering multiple days can be observed.

Similar to the other reported results in this section, there is a need for further analyzing the tweets, i.e., the message content, contained in these clusters. Furthermore, the correlation to real events happened in the corresponding regions and time spans have to be analyzed.

C. Cluster Analysis

Even though the results obtained above show that the proposed workflow enables to provide a visualization of the spatio-temporal distribution of collected twitter data related to crisis-events, there is a clear need for further analyses. In other works, it could be demonstrated, that a significant increase of the amount of tweets posted (per region and per time interval) can be a good indicator that an event occured. However, deciding wheather these anomalies are caused by virtual events, for example emerging discussions about an actual topic, or real events, is not possible. Hence, the results obtained in this work so far would be more meaningful and valuable, if the content of the clustered tweets is further analyzed.

In [28], a scalable aggregation and geolocated text visualization is proposed. In order to detect spatiotemporal clusters of term usage, the authors developed an enhanced Lloyd scheme. Furthermore, LDA is an often-used method to identify discussed topics in spatio-temporal clusters [1], [12], [13]. Before applying LDA, word frequencies were summarized for each cluster in [1]. In [24], word clouds are used for visualization. These are based on extracted keywords occuring in tweets. Finally, in [33] a thorough analysis of Twitter data recorded during three different hurricanes is provided by analyzing the distribution of information classes over time.

In this work, the distribution of occuring classes in clusters of crisis-related tweets as well as the analysis of word frequencies are applied. Following the analysis approach for crisis-related tweets in [33], we trained a CNN model proposed in [45] that classifies each crisis-related tweet into one of the seven CrisisLex classes [46]: *affected individuals*, *infrastructure and utilities*, *donations and volunteering*, *caution and advice*, *sympathy and support*, *other useful information*, and *not applicable*. The validation accuracy of this model after training with 10-fold cross validation was $\sim$ 0.65 (indpendent test F1 of DLR models submitted to TREC-IS2018: 0.55 (DLR_Simple_CNN), 0.49 (DLR_Fusion), 0.48 (DLR_Augmented)).

A per-class histogram of all crisis-related tweets in shown in figure 16. Furthermore, the corresponding percentages of occuring classes are visualized in figure 17. On the day of the landfall (September 14) a significant increase of crisis-related tweets (for example classes *sympathy and support* as well as *affected indivituals*) can be observed. After the landfall, tweets related to infrastructure and utilities are discussed more frequently and the percentage of tweets related to affected individuals has two peaks. The percentage of messages related to donations and volunteering significantly increases on September 18.

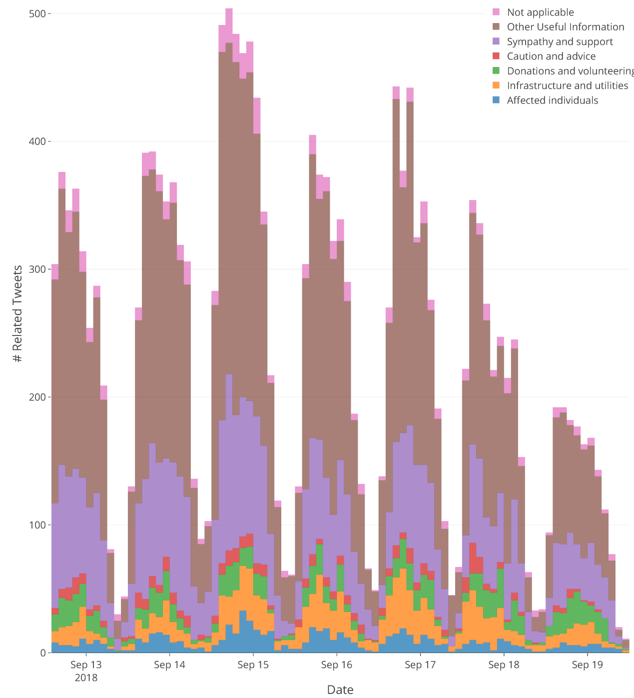

Fig. 16: Information class histogram.

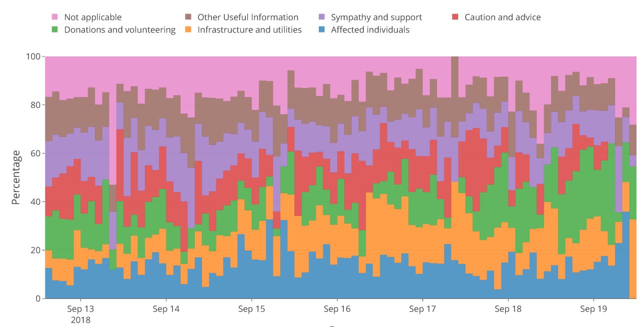

Fig. 17: Percentage of information classes.

The overall histogram of discussed topics provides a rough trend about the events happened in the affected area. Since the messages within clusters have been generated within a specific region and time, we expect that a more specific within spatio-temporal clusters. In figure 18, the ten largest clusters obtained by ST-DBSCAN (containing more than 500 tweets) are depicted. For each of these clusters, the percentage of information class occurences is visualized in figure 19. All tweets idenetified as noise by ST-DBSCAN are summarized in the cluster “Noise”, whereas the average category distribution of the ten valid clusters is denoted as “Average”. An interesting observation is that the noise and average distributions are

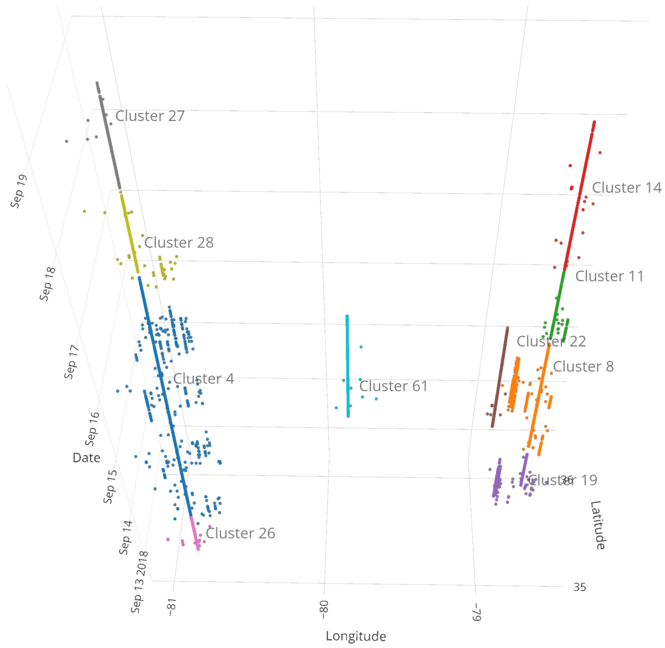

Fig. 18: Ten largest ST-DBSCAN clusters.

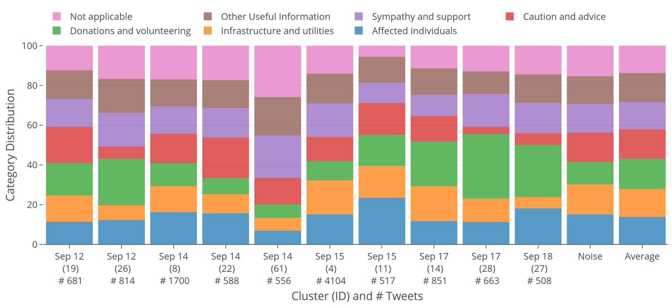

Fig. 19: Per-cluster percentage of information classes.

quite similar. This indicates that even though noise messages were posted from a low-activity twitter region, the content might still be important. Compared to the average and noise cluster distribution, the per-cluster distributions turn out to be slightly different. However, significant changes, for example the absence of one or even more of the information classes can not be observed. The distributions rather show, that in each region and during the whole period all topics related to the information classes are discussed. On the other hand, some category distribution difefrences can be observed for clusters covering the same time but generated in different regions.

In order to get a better understanding about the written content contained in each cluster, a keyword map can be generated from the raw tweet texts of a cluster. We used the Wordcloud[12] tool along with common pre-processing steps, like tokenization, lowercasing as well as removing nonalphabetic terms and stopwords. The plot of two keyword maps for clusters 4 and 14 in figures 20 and 21, respectively, further verify that the discussed topics within the clusters tend to be similar.

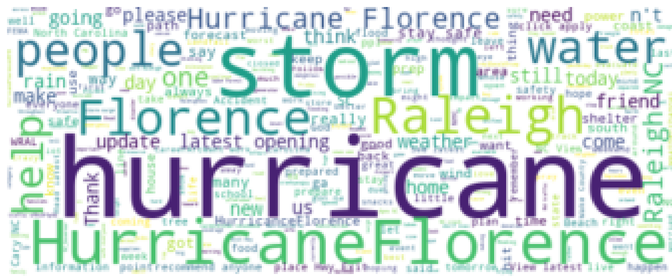

Fig. 20: Keyword map for cluster 4.

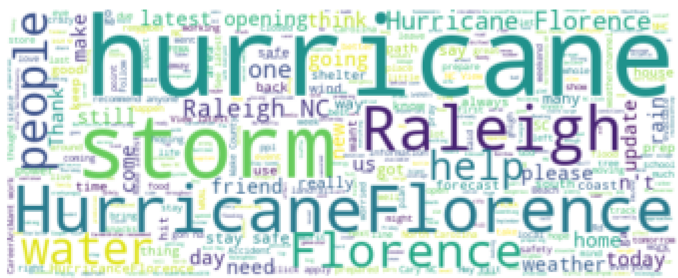

Fig. 21: Keyword map for cluster 14.

VI. Discussion

The results reported above clearly demonstrate, that spatio-temporal patterns can be identified by applying the relatively simple methods Getis-Ord G*-statistics, ST-KDE and ST-DBSCAN. However, it has to be pointed out, that the data used as input for these methods has to be pre-processed in a meaningful way. In this study, crisis-related tweets are automatically identified in the Twitter data stream recorded during hurricane Florence using a state-of-the-art CNN recently evaluated in [31]. Hence, the methods for spatio-temporal data ananlysis are applied on the subset of crisis-related tweets.

The G*-statistics are used to identify potential hot- and cold-spots in crisis-related twitter data. In order to mitigate the bias introduced by varying population densities, the amount of tweets is normalized accordingly. As a result, the area around Charlotte - the largest city of North Carolina - was identified as a cold spot, even though a similar number of crisis-related tweets were detected as in the center region of the AOI (Cary, Raleigh, Durham, Burlington, Greensboro). Hence, as expected, this statistic is a useful tool to point out high-activity regions within a specific AOI, but might neglect other high-activity regions because of a higher population density.

Even though the population density is not taken into account in STKDE, the results turn out to be meaningful for us, since only crisis-related tweets are analyzed. The overall spatio-temporal crisis-related activity densities are visualited in figures 12 and 13. The obtained density clearly reflects daily activity patterns as well as areas with high population density. A higher activity can be observed after the hurricane landfall.

[12]https://github.com/amueller/word_cloud

According to [31], the fraction of identified crisis-related tweets is around 8 % in non-crisis periods. Hence, a continuous analysis of the twitter stream would be one approach to spatio-teporally detect significant changes in crisis-relatd tweet activities.

ST-DBSCAN is a further widely-used approach, in which a tweet can either be assigned to a spatio-temporal cluster or is assumed to be noise. Our results (figures 14 and 15) show that densely populated areas tend to produce clusters covering more than one day, since tweets during the whole day and night are posted.

A critical aspect is the choice of method parameters. In this work, first results with empirically found sets of parameters are obtained. Further investigations regarding the tweet content in section V-C reveal, that the discussed topics are quite similar in all identified clusters. The choice of more appropriate parameters should therefore be investigated more systematically, in future work.

Our observation regarding the class distributions within ST-DBSCAN clusters raise the central question of **what we eventually want to capture from the data**. Screening changes in the number of crisis-related tweets posted in a region compared to a mean amount of those tweets in a non-crisis period might help to identify crisis-related activities in general. Assuming that spatial-temporally dense occurences of crisis-related tweets are caused by real-world events within that area and time, this is a quite reasonable approach. But as we saw, the topics discussed within these clusters tend to be quite similar and equally various. This means, that also the automated summarization and identification of discussed topics will probably lead to similar results for all clusters. In this regard, it might be of interest to identify all tweets related to a specific topic (e.g. *affected individuals*) from different locations and/or times rather than being posted from a similar region. In order to obtain spatial, temporal or spatio-temporal clusters related to specific sub-topics (information classes), the application of clustering methods that are able to handle both, categorial and continuous data (GMM [17], FDCA [20]) seems to be a promising approach.

A serious problem of using Twitter data is the huge amount of missing or uncertain tweet geo-locations. $\sim 99\,\%$ of all Tweets are usually not geo-located. On the other hand, this implies that fetching $1\,\%$ of all tweets roughly represents the set of all geo-located tweets [13] available via the Twitter API. However, among the geo-located tweets, there usually is a further significant fraction of tweets with a city- or state-level geo-location. In this case, the tweet coordinates are set to the corresponding centroid of the city or state, respectively. City-level geo-references are clealy visible in the ST-DBSCAN result plots (figures 14 and 15). Since the state-level geo-locations turn out to not being useful for the methods applied in this work, they are not used here. This further reduces the amount of available information and in turn introduces bias as well as increases the problems of data and information sparsity.

To overcome the sparsity issue, as many as possible available tweets - with and without geo-location - could be integrated in the analysis workflow. For instance, after the identification of spatio-temporal clusters based on the precisely geo-located tweets, the content of these clusters could be further analyzed regarding the occurence of specific keywords (see section V-C). Once, an event is detected by density analyses, like STKDE, the keywords extracted from the first tweets related to this crisis event might be used to fetch more Twitter data by keyword-based search. These tweets might then also be classified into crisis-related classes, as done in this work. Even though most of the tweets might not be geolocated, a common clustering of tweets with and without geo-location might be a good approach in order to temporarily and thematically group the tweets. The geo-located tweets can serve as spatial anchor-points for the obtained clusters. In this way, the content of tweets is integrated into clustering, leading to clusters with similar message contents.

Besides the text itself, the incorporation of additional metadata might be helpful. For instance, the user's home location, or the fact that a post is a retweet might be valuable information.

VII. Conclusion and Future Work

In this work, an initial concept for the spatio-temporal analysis of Twitter data in context of natural disasters is proposed. The goal of our work within this field is to develop concepts, methods and workflows for analyzing the impact of natural disasters on citizens as well as to gain insights that serve as a foundation for the development of further methods, like event detection and monitoring.

As a starting point, we thoroughly reviewed the current trends as well as available approaches for spatio-temporal social media data analyses. Based on this review, we constituted a first workflow, in which state-of-the-art methods for tweet filtering [31] and classification [45] of crisis-related tweets, as well as methods for spatio-temporal clustering of these tweets [10], [11], [15] are utilized. This workflow is then exemplarily applied to Twitter data recorded during hurricane Florence, in September 2018.

With a focus on methods for the spatio-temporal analysis of Twitter data, we applied local hot-spot analysis (Getis-Ord G*-statistics), spatio-temporal kernel density estimation (STKDE) as well as STDBSCAN (Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise) on the identified crisis-related tweets. The G*-statistics are a useful tool to identify hot- and cold-spots of twitter activity within a specific area. It is therefore a tool for relative analyses. Since the amount of tweets per region is normalized by the population density, comparable high activity regions might not be identified as potential hot-spots in case other regions with similar tweets amounts but lower population densities can be observed.

With ST-KDE, spatio-temporal densities of occurring crisis-related tweets can be computed and visualized. In particular, this quickly helps to identify high-activity regions and times, when tweets elated to a specific topic are taken into account.

STDBSCAN is a further approach to cluster spatio-temporally similar tweets. Those tweets not belonging to a cluster are assumed to be noise.

A critical aspect is the choice of method parameters for all three methods. In this work, first results with empirically found sets of parameters are obtained. The choice appropriate parameters should therefore be investigated more systematically, in future work.

The clusters found by STDBSCAN are further analyzed regarding the message contents. In this regard, we applied a CNN proposed in [45] to classify the crisis-related tweets into seven information classes. We then analyzed the class distributions and used keywords for each cluster. Our investigations reveal, that the discussed topics are quite similar in all identified clusters. Only minor variations providing a slight hint for the main phases of a forecasted natural disaster, i.e., preparedness, response, and recovery, can be observed. As a consequence, an automated summarization and identification of discussed topics will probably lead to similar results for all clusters.

In order to identify all tweets related to a specific topic (e.g. *affected individuals*) from different locations and/or times rather than being posted from a similar region, clustering methods that are able to handle both, categorial and continuous data (GMM [17], FDCA [20]) seem to be promising approaches for future works. The combination of different dimension, for example space and content, time and content as well as space, time and content could provide different meaningful views on the same data.

In order to mitigate the problem of missing or uncertain tweet geo-locations, more tweets without a geo-location could be retrieved by keyword-based filtering, where the keywords are derived from the geo-located tweets of the actual event. Similar to the event-related geo-located tweets, these additional tweets can then also be classified into crisis-related classes. Even though most of the tweets might not be geolocated, a common clustering of tweets with and without geo-location could be a good approach in order to temporarily and/or thematically group the tweets. The geo-located tweets can serve as spatial anchor-points for the obtained clusters. In this way, clusters with similar message contents can be obtained.

Finally, two further aspects are not covered in this work. First, the incorporation of messages not related to natural disasters could provide a hint about the usual amount of tweets posted in a region. This information might be used to find other ways of normalizing the data, for example in case of the G-statistics. Second, the validation of the workflow is a cruical step that will also be part of our subsequent work. In particular, this step will clearly help to empirically determine the method parameters for the spatio-temporal analysis.

References

[1] Y. Huang, Y. Li, and J. Shan, "Spatial-temporal event detection from geo-tagged tweets," *ISPRS Int. J. Geo-Information*, vol. 7, p. 150, 2018.

[2] E. Steiger, J. P. de Albuquerque, and A. Zipf, "An advanced systematic literature review on spatiotemporal analyses of twitter data," *Transactions in GIS*, vol. 19, no. 6, pp. 809–834, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12132

[3] C. Reuter, A. L. Hughes, and M.-A. Kaufhold, "Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research," *International Journal of Human-Computer Interaction*, vol. 34, no. 4, pp. 280–294, Apr. 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1427832

[4] G. Burel and H. Alani, "Crisis Event Extraction Service (CREES) - Automatic Detection and Classification of Crisis-related Content on Social Media," in *15th International Conference on Information Systems for Crisis Response and Management*, May 2018, p. 12.

[5] G. Burel, H. Saif, M. Fernandez, and H. Alani, "On semantics and deep learning for event detection in crisis situations," in *Workshop on Semantic Deep Learning (SemDeep), at ESWC 2017*, May 2017. [Online]. Available: http://oro.open.ac.uk/49639/

[6] D. T. Nguyen, K. A. Al-Mannai, S. Joty, H. Sajjad, M. Imran, and P. Mitra, "Robust classification of crisis-related data on social networks using convolutional neural networks," in *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*. AAAI press, 1 2017, pp. 632–635.

[7] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *International Journal of Geographical Information Science*, vol. 32, no. 1, pp. 49–72, 2018. [Online]. Available: https://doi.org/10.1080/13658816.2017.1367003

[8] L. Ngamassi, A. Malik, J. Zhang, and D. Edbert, "Social media visual analytic toolkits for disaster management: A review of the literature," in *14th Proceedings of the International Conference on Information Systems for Crisis Response and Management, Albi, France, May 21-24, 2017.*, 2017. [Online]. Available: http://idl.iscram.org/files/louisngamassi/2017/1511_LouisNgamassi_etal2017.pdf

[9] Y. Shi, M. Deng, X. Yang, Q. Liu, L. Zhao, and C.-T. Lu, "A framework for discovering evolving domain related spatio-temporal patterns in twitter," *ISPRS Int. J. Geo-Information*, vol. 5, p. 193, 2016.

[10] J. Lee, J. Gong, and S. Li, "Exploring spatiotemporal clusters based on extended kernel estimation methods," *International Journal of Geographical Information Science*, vol. 31, no. 6, pp. 1154–1177, 2017. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/13658816.2017.1287371

[11] J. K. Ord and A. Getis, "Local spatial autocorrelation statistics: Distributional issues and an application," *Geographical Analysis*, vol. 27, no. 4, pp. 286–306, 1995. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.1995.tb00912.x

[12] B. Resch, F. Uslnder, and C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," *Cartography and Geographic Information Science*, vol. 45, no. 4, pp. 362–376, 2018. [Online]. Available: https://doi.org/10.1080/15230406.2017.1356242

[13] T. Cheng and T. Wicks, "Event detection using twitter: A spatio-temporal approach," *PLOS ONE*, vol. 9, no. 6, pp. 1–10, 06 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0097807

[14] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, pp. 226–231. [Online]. Available: http://dl.acm.org/citation.cfm?id=3001460.3001507

[15] D. Birant and A. Kut, "St-dbscan: An algorithm for clustering spatial-temporal data," *Data and Knowledge Engineering*, vol. 60, no. 1, pp. 208–221, 2007, intelligent Data Mining. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169023X06000218

[16] L. Hunt and M. Jorgensen, "Clustering mixed data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 4, pp. 352–361, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.33

[17] A. Ahmad and S. S. Khan, "A survey of mixed data clustering algorithms," *CoRR*, vol. abs/1811.04364, 2018. [Online]. Available: http://arxiv.org/abs/1811.04364

[18] V. K. R. Sridhar, "Unsupervised topic modeling for short texts using distributed representations of words," in *VS@HLT-NAACL*, 2015.

[19] E. Jonsson and J. Stolee, "An evaluation of topic modelling techniques for twitter," University of Toronto, Toronto, USA, Tech. Rep., 2015.

[Online]. Available: https://www.cs.toronto.edu/~jstolee/projects/topic.pdf
[20] C. Jinyin, H. Huihao, J. Chen, Y. Shanqing, and S. Zhaoxia, “Fast density clustering algorithm for numerical data and categorical data,” *Mathematical Problems in Engineering*, vol. 2017, pp. 1–15, 03 2017.
[21] D. Ramachandran and P. Ramasubramanian, “Event detection from twitter - a survey,” *International Journal of Web Information Systems*, vol. 14, no. 3, pp. 262–280, 2018. [Online]. Available: https://doi.org/10.1108/IJWIS-11-2017-0075
[22] M. Hasan, M. A. Orgun, and R. Schwitter, “A survey on real-time event detection from the twitter data stream,” *Journal of Information Science*, vol. 44, no. 4, pp. 443–463, 2018. [Online]. Available: https://doi.org/10.1177/0165551517698564
[23] M. Walther and M. Kaisser, “Geo-spatial event detection in the twitter stream,” in *Advances in Information Retrieval*, P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, I. Segalovich, and E. Yilmaz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 356–367.
[24] X. Zhou and C. Xu, “Tracing the spatial-temporal evolution of events based on social media data,” *ISPRS International Journal of Geo-Information*, vol. 6, no. 3, 2017. [Online]. Available: http://www.mdpi.com/2220-9964/6/3/88
[25] F. Santa, R. Henriques, J. Torres-Sospedra, and E. Pebesma, “A statistical approach for studying the spatio-temporal distribution of geolocated tweets in urban environments,” *Sustainability*, vol. 11, no. 3, 2019. [Online]. Available: http://www.mdpi.com/2071-1050/11/3/595
[26] Liang, Chunyang, Lin, Guangfa, and Peng, Junchao, “Discover the spatio-temporal process of typhoon disaster using micro blog data,” *MATEC Web Conf.*, vol. 246, p. 03013, 2018. [Online]. Available: https://doi.org/10.1051/matecconf/201824603013
[27] Y. Martn, Z. Li, and S. L. Cutter, “Leveraging twitter to gauge evacuation compliance: Spatiotemporal analysis of hurricane matthew,” *PLOS ONE*, vol. 12, no. 7, pp. 1–22, 07 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0181701
[28] D. Thom, H. Bosch, S. Koch, M. Wrner, and T. Ertl, “Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages,” in *2012 IEEE Pacific Visualization Symposium*, Feb 2012, pp. 41–48.
[29] Y. Hu, F. Wang, C. Guin, and H. Zhu, “A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation,” *Applied Geography*, vol. 99, pp. 89 – 97, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0143622818300560
[30] L. Zhao, F. Chen, J. Dai, T. Hua, C.-T. Lu, and N. Ramakrishnan, “Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling,” *PLOS ONE*, vol. 9, no. 10, pp. 1–12, 10 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0110206
[31] J. Kersten, A. Kruspe, M. Wiegmann, and F. Klan, “Robust filtering of crisis-related tweets,” in *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, Valencia, Spain, May 19-22, 2019*, 2019, accepted, to be published.
[32] A. Kruspe, J. Kersten, and F. Klan, “Detecting event-related tweets by example using few-shot models,” in *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, Valencia, Spain, May 19-22, 2019*, 2019, accepted, to be published.
[33] F. Alam, F. Ofli, and M. Imran, “CrisisMMD: Multimodal Twitter Datasets from Natural Disasters,” *arXiv:1805.00713 [cs]*, May 2018, arXiv: 1805.00713. [Online]. Available: http://arxiv.org/abs/1805.00713
[34] R. McCreadie and I. Soboroff, “Trec overview paper - incidents streams track,” in *26th Text Retrieval Conference*, 2018, p. 9.
[35] J. Kersten, “Simultaneous feature selection and gaussian mixture model estimation for supervised classification problems,” *Pattern Recognition*, vol. 47, no. 8, pp. 2582 – 2595, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320314000673
[36] P. C. Cheeseman and J. C. Stutz, “Bayesian classification (autoclass): Theory and results,” in *Advances in Knowledge Discovery and Data Mining*, 1996.
[37] C. S. Wallace and D. L. Dowe, “Mml clustering of multi-state, poisson, von mises circular and gaussian distributions,” *Statistics and Computing*, vol. 10, no. 1, pp. 73–83, Jan 2000. [Online]. Available: https://doi.org/10.1023/A:1008992619036
[38] M. A. Jorgensen and L. A. Hunt, “Mixture model clustering of data sets with categorical and continuous variables,” in *Information Statistics and Induction in Science*, D. L. Dowe, K. B. Korb, and O. J. J, Eds., 1996.
[39] L. Hunt and M. Jorgensen, “Mixture model clustering using the MULTIMIX program.” *Aust. N. Z. J. Stat.*, vol. 41, no. 2, pp. 153–171, 1999.
[40] V. Epanechnikov, “Non-parametric estimation of a multivariate probability density,” *Theory of Probability & its Applications*, vol. 14, no. 1, pp. 153–158, 1969. [Online]. Available: https://doi.org/10.1137/1114019
[41] Google Maps, “V 7.3.2.5495, north carolina, usa, lat 35.25, lon -78.91, eye alt 705 feet. sio, noaa, u.s. navy, nga, gebco. landsat, copernicus, google 2018, accessed at december 19, 2018,” 2018.
[42] NASA Earth Observatory, https://earthobservatory.nasa.gov/images/92775/florence-inundates-the-carolinas, accessed: December 19, 2018.
[43] ESRI, https://www.arcgis.com/home/item.html?id=ff48bbae433442a38f6c635b8c7baf72, 2016 Population Density by Congressional District, accessed: March 3, 2019.
[44] Z. Wang, X. Ye, and M.-H. Tsou, “Spatial, temporal, and content analysis of twitter for wildfire hazards,” *Natural Hazards*, vol. 83, no. 1, pp. 523–540, Aug 2016. [Online]. Available: https://doi.org/10.1007/s11069-016-2329-6
[45] A. Kruspe, J. Kersten, M. Wiegmann, B. Stein, and F. Klan, “Classification of incident-related tweets: Tackling imbalanced training data using hybrid cnns and translation-based data augmentation,” in *Proceedings of the The Twenty-Seventh Text REtrieval Conference (TREC 2018)*, 2018.
[46] A. Olteanu, S. Vieweg, and C. Castillo, “What to expect when the unexpected happens: Social media communications across crises,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*.