Analyzing Refugee Migration Patterns Using Geo-Tagged Tweets

Franziska Hübl [1], **Sreten Cvetojevic** [2], **Hartwig Hochmair** [2] (iD) **and Gernot Paulus** [1,*] (iD)

[1] Geoinformation and Environmental Technology, Carinthia University of Applied Sciences, Villach 9524, Austria; franziska.huebl@alumni.fh-kaernten.at

[2] Geomatics Program, Fort Lauderdale Research and Education Center, University of Florida, Davie, FL 33314, USA; scvetojevic@ufl.edu (S.C.); hhhochmair@ufl.edu (H.H.)

[*] Correspondence: g.paulus@fh-kaernten.at

Abstract: Over the past few years, analysts have begun to materialize the "Citizen as Sensors" principle by analyzing human movements, trends and opinions, as well as the occurrence of events from tweets. This study aims to use geo-tagged tweets to identify and visualize refugee migration patterns from the Middle East and Northern Africa to Europe during the initial surge of refugees aiming for Europe in 2015, which was caused by war and political and economic instability in those regions. The focus of this study is on exploratory data analysis, which includes refugee trajectory extraction and aggregation as well as the detection of topical clusters along migration routes using the V-Analytics toolkit. Results suggest that only few refugees use Twitter, limiting the number of extracted travel trajectories to Europe. Iterative exploration of filter parameters, dynamic result mapping, and content analysis were essential for the refinement of trajectory extraction and cluster detection. Whereas trajectory extraction suffers from data scarcity, hashtag-based topical clustering draws a clearer picture about general refugee routes and is able to find geographic areas of high tweet activities on refugee related topics. Identified spatio-temporal clusters can complement migration flow data published by international authorities, which typically come at the aggregated (e.g., national) level. The paper concludes with suggestions to address the scarcity of geo-tagged tweets in order to obtain more detailed results on refugee migration patterns.

Keywords: tweets; trajectory; migration; hashtag; refugee; V-Analytics

1. Introduction

Since the beginning of 2015, Europe has been experiencing the largest migration movement in the 21st Century, which is caused by political crises and wars in the Middle East and Africa. Figure 1 provides a snapshot of refugee movements to Europe in July 2015 based on country level data that are published by the UN Refugee Agency (UNHCR) and visualized by Saarinen and Ojala [1]. The flow intensity on the map, where each moving dot represents 25 people, shows that Syria, Afghanistan and Iraq are the countries where the majority of refugees originate from. White vertical bars indicate the numbers of asylum seekers per destination country between 2012 and July 2015. Germany can be identified as the most prominent refugee destination country in the world, with a total of 441,900 asylum seekers registered during 2015 [2]. Although migration routes are shown in their entirety, they are computed as shortest paths connecting country centroids and therefore do not reflect actual migration corridors taken. It is also important to note that UNHCR numbers rely on registered refugees, which will deviate from actual refugee arrival numbers. Reasons can be delays in the governmental registration process, some nations not releasing all of their refugee statistics, or discrepancies between countries on how to count appeals to asylum denials. In addition,

the European Union (EU) Dublin regulation states that, in whichever country a refugee is first registered and has his or her fingerprints taken, that country is to handle the asylum application, that is, to accept or reject the claim. Many migrants know about this rule, some of which deliberately try to avoid registration in their country of first entry to the EU (e.g., Greece) since such prior registration would reduce their chance of being allowed to remain in richer countries of the EU (e.g., Germany or Sweden) if they already moved there after entering EU territory [3]. Also, falsely stated information about a refugee's home country to increase the chance for political asylum in Europe may lead to additional data bias in official datasets. For example, Germany estimates that 30% of incoming migrants who claim to be Syrian citizens are from other countries [4]. Though based on a reliable data source, the map content of Figure 1 could be misused by political parties or individuals to back their agenda, e.g., as a proof of the alleged threat of refugees to Europe, if taken out of context. Numerous examples of historic maps exist that were purposely used for political propaganda through manipulation of map design (e.g., choice of map projection, colors, line style, and size of map symbols) [5–7]. In Figure 1, for example, the size and density of arrows could be used to influence the impression that Europe is, or is not, experiencing a heavy inflow of refugees.
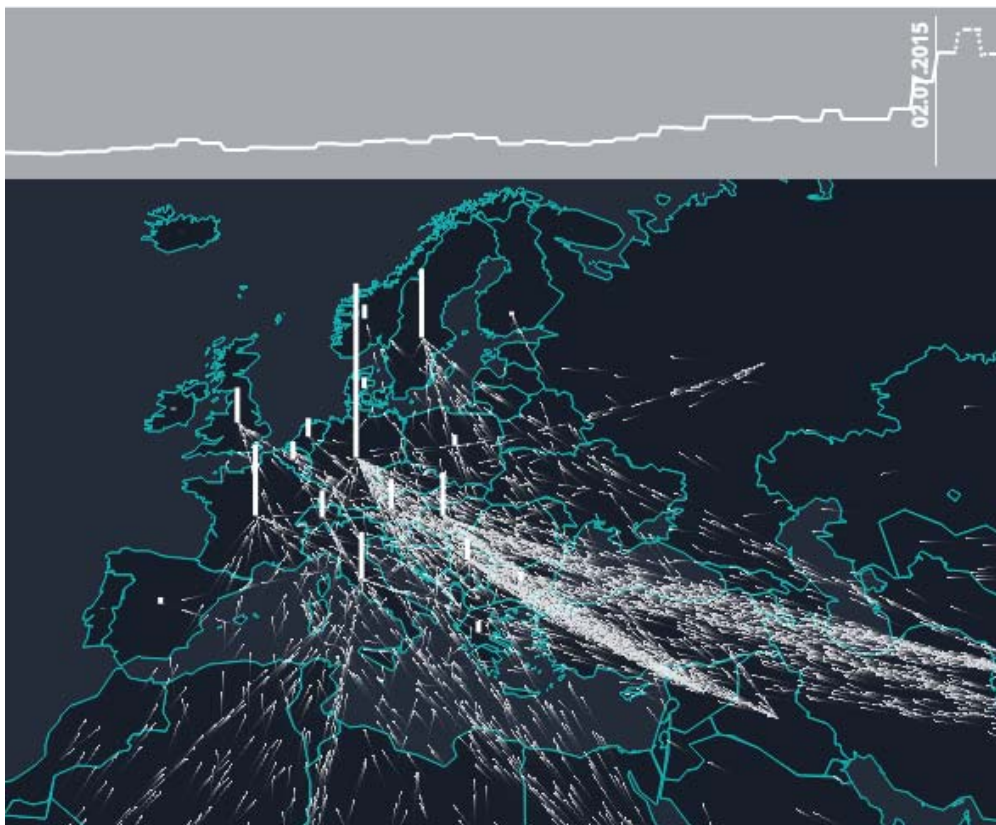

Figure 1. Refugee flows from Africa and the Middle East to Europe in July 2015 with a histogram (on top) showing the daily numbers of asylum seekers (adapted from [1]).

Understanding migration patterns is important for administrative logistics such as accommodation, transportation, education, and distribution of refugees as well as for the mitigation of refugee causes. Social media platforms and phone records are potential data sources for studying refugee movement patterns besides more traditional data sources, such as governmental datasets or statistical data from national or international agencies or voluntary aid programs. However, especially for crowd-sourced and social media data, such as tweets, it must be noted that these data undergo a selection bias and do not necessarily represent the entire population [8,9]. Twitter is a microblogging service that allows its users to send 140 character long posts called tweets. It is utilized every month

by over 300 million active users, and therefore frequently used to study communication patterns and information flows among people [10,11]. Drawing from geo-positioning capabilities of mobile devices through GPS, Wi-Fi, or cell phone towers, tweeted posts can be annotated with location information about the user's current position if the user opts to do so in the application. However, only about 1% of all tweets is geo-tagged [12], which means that results of any analysis that involves Twitter geo-locations are not necessarily representative of all Twitter users. Geo-tagged positional information within tweets can be relayed as point with exact geographic coordinates or as bounding box when the location is expressed as a place, which is less precise. Despite the small percentage of tweets that carry positional information (either as exact coordinates or as bounding box), tweets have been used for the analysis of travel patterns at the local [13], national [14,15] and global [16,17] scale, and for the detection and tracking of spatial events [18]. Being restrained by the scarcity of geo-tagged tweets in low-density population areas and in countries with low Twitter penetration rates, most studies analyze travel behavior in metropolitan areas, base their movement analysis on aggregated user flows between large areal units (e.g., counties, states, or countries), or provide statistical summaries about movements (e.g., peak times and distances) rather than visualizing individual movement trajectories or trajectory clusters in maps. Being aware of low Twitter penetration rates in many countries of the Middle East and Africa [17], we expand previous work by exploring means to use tweets for the characterization of refugee migration patterns. More specifically, the presented study evolves around the following two objectives:

- Extract individual and aggregated trajectories that reflect refugee migration movements from crisis-ridden countries in the Middle East and Africa to Europe; and
- Identify spatio-temporal event clusters of refugee related tweets to determine likely locations of refugees along migration routes and areas of elevated tweet activities related to refugees in destination countries.

The second objective is independent of trajectory analysis but instead uses hashtag filters to identify clusters of refugee mentions in local communities. The proportion of geo-tagged refugee related tweets among all geo-tagged tweets in administrative units (e.g., provinces) of destination countries (e.g., Germany) can identify areas that are highly concerned with migration topics.

The overarching goal of this study is to identify spatial and spatio-temporal patterns of refugee movements based on geo-tagged tweets. In this context, the two above-mentioned objectives follow a framework that involves the view of community-based, space-time referenced data in a dual way, which includes: (1) trajectories of people; and (2) independent spatio-temporal events [19]. Trajectory analysis requires a high quantity of observable routes taken by individuals. Extracted trajectories can be subsequently analyzed within agent-centered tasks in order to describe properties and behaviors of people (in general, moving agents). Trajectories can also be clustered and visualized as flow maps, naturally expressing the dynamic aspect of movement data. The dependency on trajectories from individuals for this type of analysis, though, limits the information that can be extracted in the case of sparse data, as will be showcased with the Twitter dataset at hand. Spatio-temporal events, on the other side, are defined through spatio-temporal positions and thematic attributes. Identification of such events is not constrained by the search for individual moving objects, but rather defined by properties of space and places. Space-centered tasks can therefore lead to a more comprehensive, albeit different, description of the spatio-temporal phenomenon in question, e.g., refugee movements. This study will utilize both types of analyses and discuss outcomes, strengths and weaknesses of each approach.

The remainder of the paper is structured as follows: Section 2 reviews existing frameworks of movement analysis and previous work on the use of crowd-sourced data, particularly tweets, for mobility and migration analysis. Section 3 describes data extraction and management methods within the PostgreSQL data base environment as well as utilized aggregation and visualization tools in the V-Analytics software package. This is followed by Section 4, which explains exploratory methods to extract refugee trajectories. Section 5 focuses on the detection of refugee related tweet clusters through hashtag filtering. Section 6 discusses the obtained results and provides directions for future work.

2. Related Work

Movement is a central aspect of human everyday life and its understanding is therefore important for informed decision making in domains such as transportation, urban planning, tourism, housing, or finances. Movement analysis involves various types of objects, including spatial objects (with a position given in space), spatial events (defined through spatio-temporal positions and thematic attributes), moving objects (following a trajectory), and moving events (events with their spatial position changing over time), resulting in complementary views of movement including mover-oriented perspective, space-oriented perspective, time-oriented perspective, and event-oriented perspective [20]. Accordingly, movement can be represented in different types of data, including trajectories, spatial event data, local time series, and spatial distributions.

Common movement analysis techniques include space-centered tasks to study properties of the space and places (e.g., find places of interest and investigate flows between places), and agent-centered tasks (e.g., discover meetings of people and discover routes frequently taken) [19]. Other previous work distinguishes between three main categories of movement data analysis, which describe: (a) movement characteristics of an individual within a group; (b) the dynamics of a given group (e.g., direction, speed, change in group size and shape); and (c) differences between the behaviors of multiple groups (e.g., relative movements) [21].

Positioning capabilities of mobile devices through GPS, WiFi and other technologies facilitated the emergence of crowd-sourced data collections that provide a rich repository of geo-tagged and time-stamped point or trajectory information. This is oftentimes used as data basis for movement analysis and activity detection. Examples of these data sources include sport and fitness apps, such as Endomondo or Strava [22]; photo-sharing applications, such as Flickr or Panoramio [23–25]; business apps, such as Foursquare/Swarm [26]; or social media platforms, such as Twitter [17,27]. Tweets contain besides the message itself a rich set of metadata about the tweet (e.g., creation date) and the user (e.g., set language). A review of 92 papers that analyze Twitter usage patterns showed that 33% of these papers utilize all information layers (spatio-temporal and semantic), that 10% rely on spatio-temporal information only, and that the remaining 57% analyze only the semantic information (e.g., follower and following activities, hashtags, and user profiles) [28]. One study used both spatio-temporal and semantic tweet information to accurately derive trajectories of events (e.g., concerts) and to characterize these trajectories by looking into changes of sentiments over time using drift analysis [29]. The approach filters tweets through keywords and generates maps using Kernel Density Estimation (KDE) over time, which allows deriving spatial trajectories of hot spots. Sentiment and topic drift analysis are then applied to see how the user perception of these episodic hotspots changes over time, e.g., before and after the cancellation of a concert. The paper analyzed so called indirectly observed trajectories which have an implicit spatial reference and can be extracted by keyword filtering over a set of geo-tagged tweets, e.g., in the case of extracting the movement of a flood. Hashtags, keywords, and other message content of geo-tagged tweets were also used in numerous other studies to identify the spatio-temporal dimension and dynamics of local events, such as sport games, floods, earthquakes, or terrorist attacks [30–32]. For example, to identify the location of an earthquake center and a trajectory center of a typhoon, one study applied in a first step a support vector machine that, based on keywords, the number of words, and the context (i.e., the word before and after the keyword), classified a tweet as event related or not [33]. In a next step, a probabilistic spatio-temporal model that applies Kalman and particle filtering estimated the center and trajectory of the event of interest.

Despite the wide use of tweets in movement and event analysis, only few examples in the literature use tweets to identify refugee and migration patterns. One study used geo-tagged data from about 500,000 Twitter users between 2011 and 2013 to infer trends of out-migration rates for various countries around the world [34]. Results reveal a decline in out-migration rates from Mexico to other countries, but a continued relative increase in out-migration rates for Spain, Greece, and Ireland, which were hit hard by the bad economy at that time. Using a cross-national dataset of countries of

asylum and countries of origin from the Office of the United Nations High Commissioner for Refugees (UNHCR) and the United Nations Relief and Works Agency (UNRWA), Rüegger and Bohnet [35] examined refugee flight patterns from over three decades. The study found that many ethnic groups flee to countries with similar ethnic refugee groups, which complements earlier studies that identify geographical proximity as the most important factor for flight destinations [36]. As opposed to movement analysis itself, sentiment analysis about refugees and migration movements is more frequently discussed in the literature. For example, Rettberg and Gajjala [37] examined images and words shared on the Twitter hashtag #refugeesNOTwelcome to understand the portrayal of male Syrian refugees in a post-9/11 context. Another study analyzed the immediate response to Paris attacks towards Middle Eastern refugees in Europe and reviewed main topics of supportive tweets (e.g., disassociating refugees from terrorism) and anti-refugee tweets (e.g., tying refugees to the Paris attacks) [38].

3. Data and Methods

3.1. Data Extraction

Worldwide geo-tagged tweets posted between 15 October 2014 and 2 December 2015 were downloaded in JavaScript Object Notation (JSON) format through the Twitter Streaming Application Programming Interface (API) using the Tweepy python library [39]. To stay within the download bandwidth, data were downloaded separately for seven world regions and then stored in a PostgreSQL database. The two tables that covered the geographic areas of interest (enclosing most of Europe, Asia and Africa north the equator) and the selected time frame (between 2014 and 2015) contained a total of 956,689,311 geo-tagged tweets. The JSON file that comes with each tweet contains metadata about the user (e.g., user id, creation date of account, and number of followers) and the tweet itself (e.g., tweet id, text, creation date and time, place, bounding box, place type, position coordinates, and source) (Figure 2). Creation date and time are given in UTC, which provides a common time frame for all time zones involved in movement and cluster analysis.
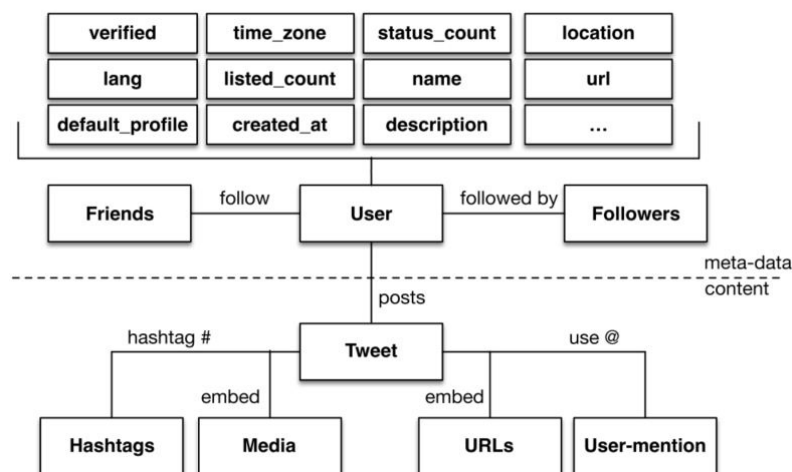

Figure 2. Twitter infographic (adapted from [40]).

Both tweets that were geo-tagged with exact coordinates or that were geo-tagged with a place description were included in the analysis. For the latter type, bounding boxes associated with Twitter places can extend over continents. Therefore, for the purpose and scale of the presented study (refugee migration patterns between countries and cities), tweets that had a polygon size over 15,000 km^2 (which corresponds to the size of the world's largest metropolitan areas) or that were of place type "Country" were removed from further analysis. Remaining polygons were subsequently replaced by their centroids and represented as point geometries. Filtered and modified tweet tables from the

different regions were finally merged into one tweet table containing tweet id, user id, tweet text (including hashtags), point geometry, and time stamp of all tweets. From this table, a user table was extracted which stored for all users the ids of their posted tweets. A query joining both tables could then be used to extract for each user id a chronologically ordered list of geo-tagged tweets. Figure 3 visualizes the sequence of further filtering steps applied to those two tables. In a first step, tweets were clipped to Europe, parts of Central Asia, Central and Northern Africa, and the Middle East (Figure 4). That whole extent was used for trajectory movement analysis, whereas only a subset of countries within that selected region (i.e., Greece, Germany, Austria, and Italy) were used for hashtag-based cluster analysis. Since this research concentrates on refugees, which are most likely moving by ground based transportation in vehicles, e.g., cars or busses, or by foot, a speed filter of 200 km/h between consecutive tweets of a user was applied. Tweets of users whose travel speed exceeded this threshold were removed from further analysis. This speed filter was combined with the exclusion of users with more than 150 tweets per day, which is indicative of automated bots [41]. Further data selection steps depended on the research objective under consideration, as detailed by the two branches in Figure 3.
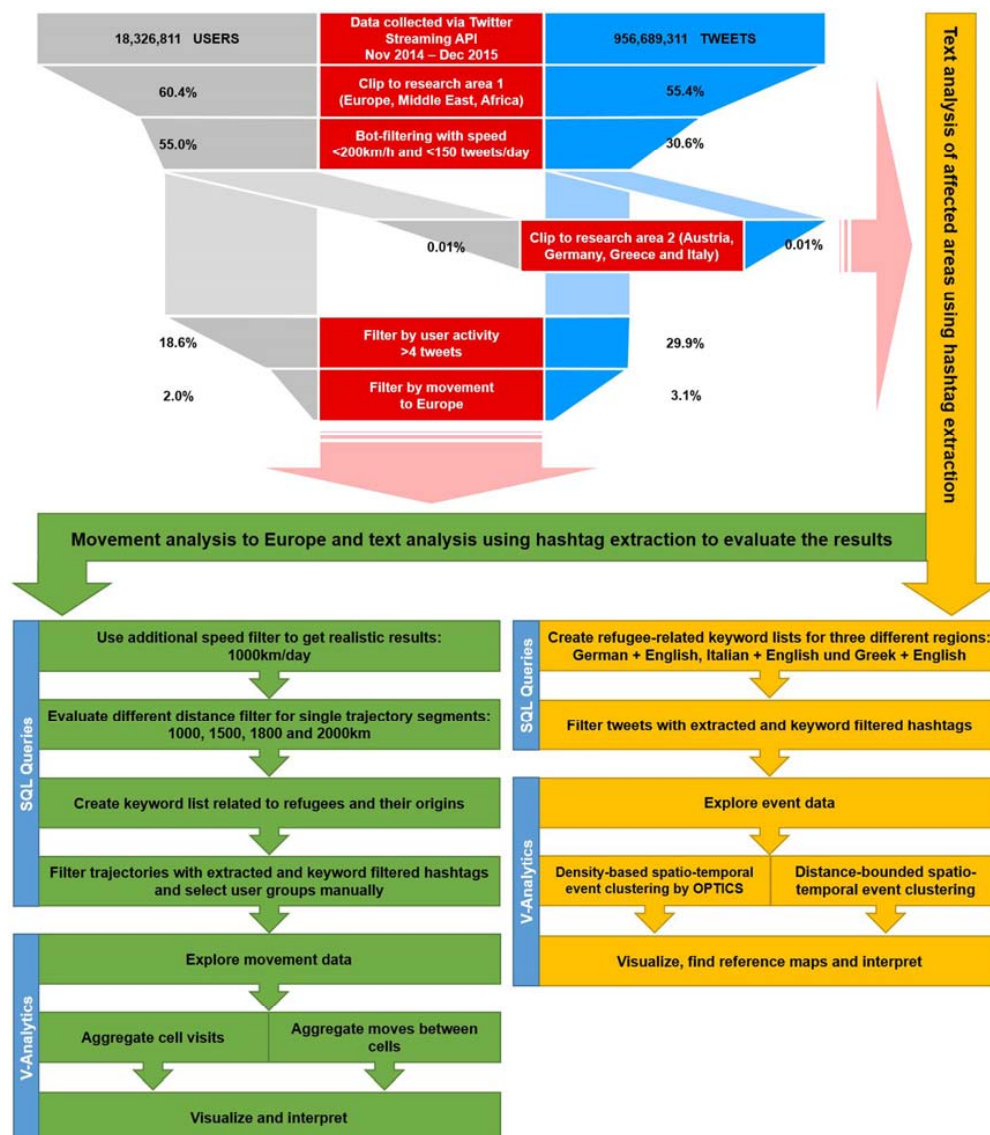

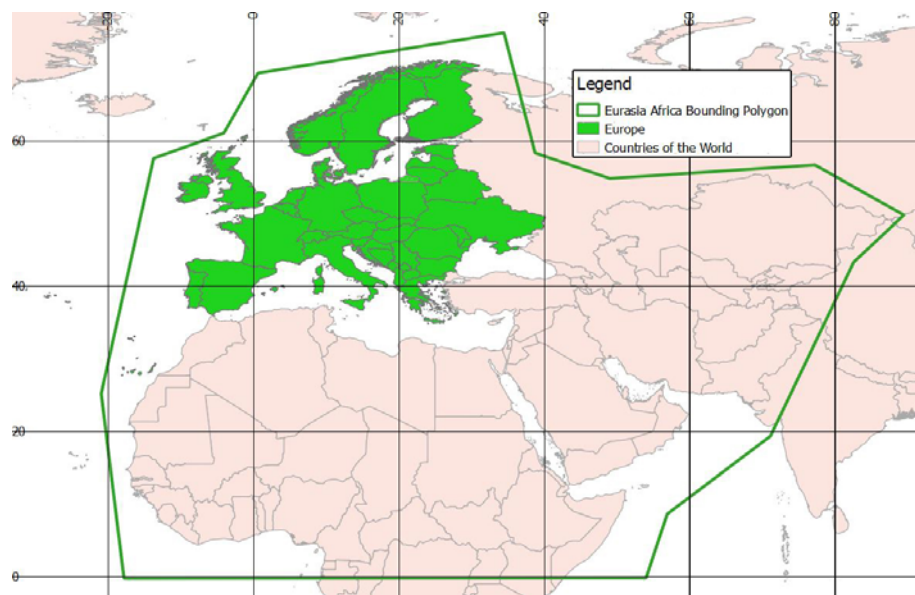**Figure 3.** Conceptual analysis model of data filtering and processing.

Figure 4. Spatial filter applied to tweets.

3.2. Methodology

Different methods were applied for movement analysis and hashtag cluster analysis. Movement analysis extracts trajectories through the application of speed, distance, and hashtag filters, and then uses exploratory tools in V-Analytics on these tweet trajectories to aggregate and generalize them as moves between existing areas [42]. As part of movement analysis, the textual content of individual trajectories will be analyzed as well. The output of the process is then used to construct a classification of users into different user groups.

Hashtag based filtering, independent of individual user trajectory analysis, is used to identify space-time clusters of refugee related tweets (primarily within the local population) in countries along the refugee routes and in destination countries. Clusters will be identified through density-based and distance-bounded spatio-temporal event clustering procedures in V-Analytics and be closely analyzed for four European countries. The trend in refugee related tweet numbers in these selected countries over time can then be compared to UNHCR estimates of asylum seekers as a method of data validation.

4. Trajectory Based Movement Analysis

4.1. Filtering Datasets

To find trajectories potentially reflecting refugee migration flows to Europe the tweets of a user under consideration need to have at least one location outside of Europe followed by a location in Europe (compare Figure 3). As opposed to this, users whose tweet patterns reflect movement from Europe to a non-European country were excluded. The spatial extent that is shown in Figure 4 was used for the trajectory based movement analysis.

4.2. Extracting Movement Patterns

The V-Analytics software, which is based on CommonGIS, generalizes and aggregates movement data, and therefore transforms trajectories into aggregate flows between areas. Trajectories consist of sorted lists of points with time stamps. Downloaded geo-tagged tweets with their geographic coordinates and time stamps (in UTC) provide the necessary information to form trajectories, which can be filtered through SQL queries (see Figure 3, left) before data aggregation. A detailed description

of the algorithm for the generalization of movement data in V-Analytics can be found in [14,43]. The aggregation algorithm in V-Analytics comprises the following steps:

1. Extract characteristic trajectory points

    These include start and end points, points of significant turns and points of significant stops (i.e., pauses in the movement).

2. Group extracted points by spatial proximity

    The implemented cluster algorithm is capable of producing convex spatial clusters with desired spatial extents. The desired radius of a group needs to be provided as a parameter.

3. Partition the area

    The study area is subdivided into Voronoi cells using centroids of groups found in Step 2 and additional points that are generated in a regular manner if they are more than twice the desired radius away from centroid points. The Voronoi cells are used as the locations for aggregating movement data and building flow maps.

4. Divide trajectories into segments

    A place-based division is devised where a trajectory is represented as a sequence of Voronoi cells generated in Step 3. A trajectory is stored as a dual representation, namely as a sequence of cell visits and a sequence of moves between cells.

5. Data aggregation

    Data are aggregated in two complementary ways. First, visits to each cell are aggregated in the form of counts, statistics of durations, path lengths, etc. Second, moves between cell pairs are aggregated and for each aggregation, some summary statistics (e.g., number of elementary moves, statistics of lengths) are computed. Aggregated moves can then be represented by arrows with width proportional to the counts of elementary moves, upon which temporal, spatial, or attribute filters can be applied.

    Using a set of tweet sequences that already underwent regional, speed (bot), activity, and directional filters as described in Section 3.1, Figure 5 shows a generalized flow map of Twitter based trajectories that have at least one stop in Germany or Austria. Occasional arrows connecting other countries (e.g., the Netherlands and France) can also be found since the entire travel route of users crossing Germany or Austria is shown. Different aggregation filter settings were explored, using settings from other studies as guidelines [42]. A visually appealing map was achieved by allowing aggregate moves at the regional level (with segment lengths between 10 and 1000 km) and setting a minimum threshold of 10 identical moves between two cells from different trajectories to avoid cluttering. More specifically, following parameter settings were used to generate the map:

    - Minimum number of trajectories per arrow = 10
    - Maximum number of trajectories per arrow = 250
    - Minimum angle between consecutive trajectory segments to be considered as a significant turn = 30 degrees
    - Minimum distance to next position = 10 km
    - Maximum distance to next position = 1000 km

As the map shows, aggregated trajectories from outside Europe originate exclusively from Turkey, which does not portray the typical pattern of refugee movements that actually originate largely from Syria and other countries of the Middle East. Therefore, additional trajectory filters, based on SQL queries, were applied as shown along the left workflow in Figure 3 (green boxes). This includes an additional speed filter of a maximum of 1000 km/day, which was assumed to be realistic for

refugee movements that are often interrupted by changes in transportation modes, stops in refugee camps or stops at borders along their travel. Although we expected to see detailed trajectories with stops in countries along expected refugee routes, still no trips from countries in the Middle East were returned. Therefore, the maximum distance to the next position was step-wise increased from 1000 km to 1500 km, 1800 km, and 2000 km, respectively. The last setting resulted in several aggregated trajectories originating from Syria, Iraq, Pakistan, or Chad (as will be discussed in the next section), since it allowed trajectories with less detailed geometries of stop points (and hence less information about their exact route) to be retained.
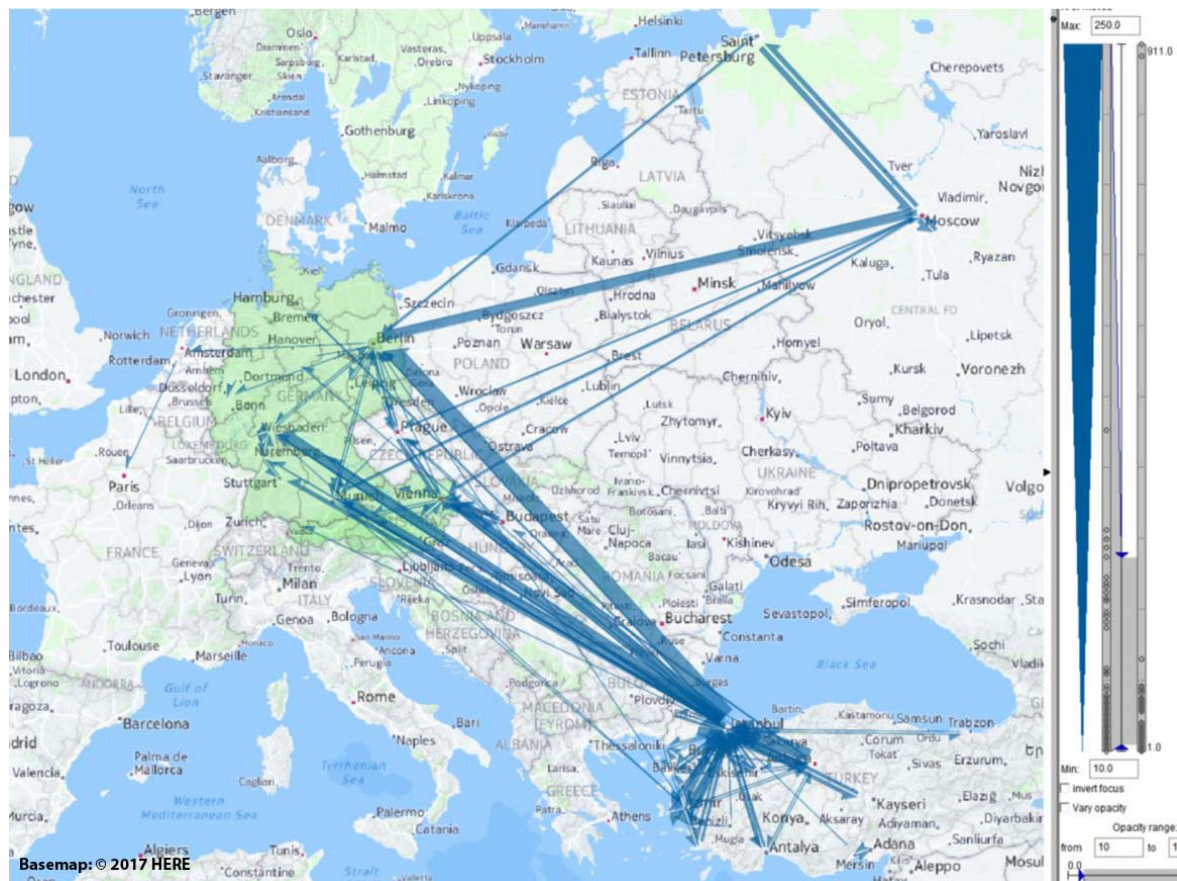

Figure 5. Generalized trajectories with at least one stop in Austria or Germany.

4.3. User Types and Movement Patterns

In addition to the previously described areal, speed, distance, and count filters tweets were filtered based on hashtags (i.e., a string of characters preceded by the hash (#) character). Hashtags are suitable to search for tweets of certain topics since they are generated by users as a method to categorize content and to highlight topics, promoting folksonomy [44]. For this purpose, a keyword list was first created that consisted of English words and word stems relating to refugee migration topics (e.g., "refugee", "asylum", and "camp"), refugee home countries (e.g., "syria", "iraq", and "afghanistan") or political topics in the Middle East (e.g., "assad" and "isis"). This list was populated in an exploratory manner by parsing through prominent tweets and various news outlets reporting about refugees. Next, all tweets that matched the keyword list were extracted, and from these tweets a list of hashtags was extracted that matched words or word stems from the keyword list. All identified hashtags were then checked manually and those not related to the topics above were removed. Next, tweets from those users who posted at least one tweet with a hashtag from the filtered hashtag list were retained for further analysis

and their trajectories extracted. A total of 5534 tweets from 37 users remained after this hashtag based filter process.

Based on the tweet content the users who posted these tweets were manually classified into different user types. Whereas manual classification of content is possible for a small number of tweets, automated procedures, including machine learning techniques (e.g., Decision Trees or Support Vector Machines) should be used for a larger set instead [33,45]. Those automated approaches would require the preparation of a manually annotated training set of tweets (indicating their user type in this example) and the selection of features that potentially help to predict the user type, such as the number of words in a tweets message, keywords in a tweet, or the presence of URLs. Figure 6 shows the proportion of tweets falling into the different user categories as found with the manual approach. Six users (16.2%) are potential refugees. Their tweet content is mainly in Arabic and English, and hashtags appear to relate to refugee concerns rather than typical tourist topics. These tweets include hashtags, such as #Assad, #Swaida, #NasserAllah, #Syria, #Israel, #Isis, #France, #world, and #Prayers4Paris. Although content in Arabic was not included in the text analysis, based on the used English hashtags, most of these tweets express concerns about the political situation in the Middle East.
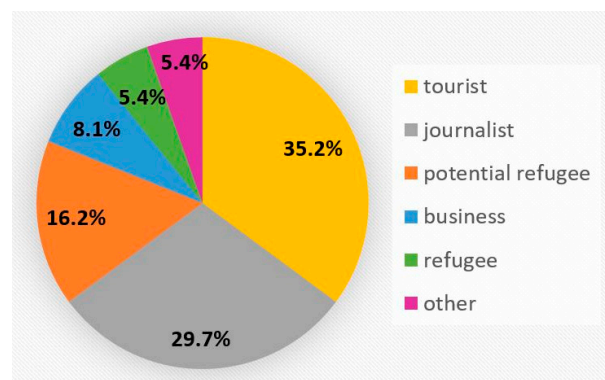

Figure 6. User groups identified through manual content analysis.

Two users (5.4%) can be classified as refugees with high certainty. One of them stated to be a Syrian refugee and tweeted about the conditions and events in their refugee camp. The other user describes his life situation as a refugee from Lebanon. Both users report about their German lessons and one of them even started to post German tweets.

Figure 7 visualizes trajectories of the 37 manually classified users in V-Analytics with generalized and summarized trajectories grouped per user type, using a maximum distance threshold of 2000 km between two cell areas. The maximum number of five users per trajectory indicates a very low number of identified people given that the data source covers one year of tweets. This data scarcity illustrates the previously mentioned major limitation of agent-centered tasks with social media in the context of refugee pattern analysis.

Figure 8 shows the individual trajectories of the two identified likely refugees as blue lines (time runs along the z-coordinate axis) and their aggregated routes as green arrows. These two users did not tweet along their entire trip from the Middle East to Germany. Besides the lack of waypoints, their tweet text does not provide clues about their chosen route either. Only their home countries Lebanon and Syria and their destination country (Germany) are mentioned. However, using geo-tagged tweets a detailed trajectory of their routes within Germany can be produced (which is not mapped here to maintain user privacy). It is also likely that the refugees applied for asylum since their trajectory waypoints are located near migration centers and other refugee related institutions in Germany.
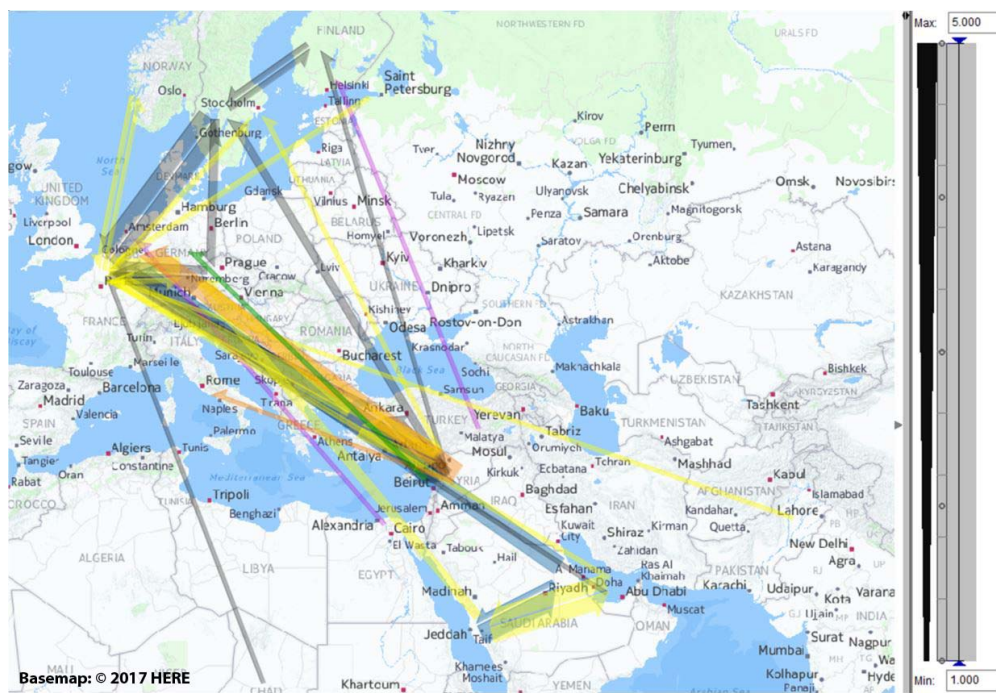
Figure 7. Aggregated routes to Europe for different user types: business (blue), journalist (dark grey), tourist (yellow), potential refugee (orange), likely refugee (green), and other (magenta).
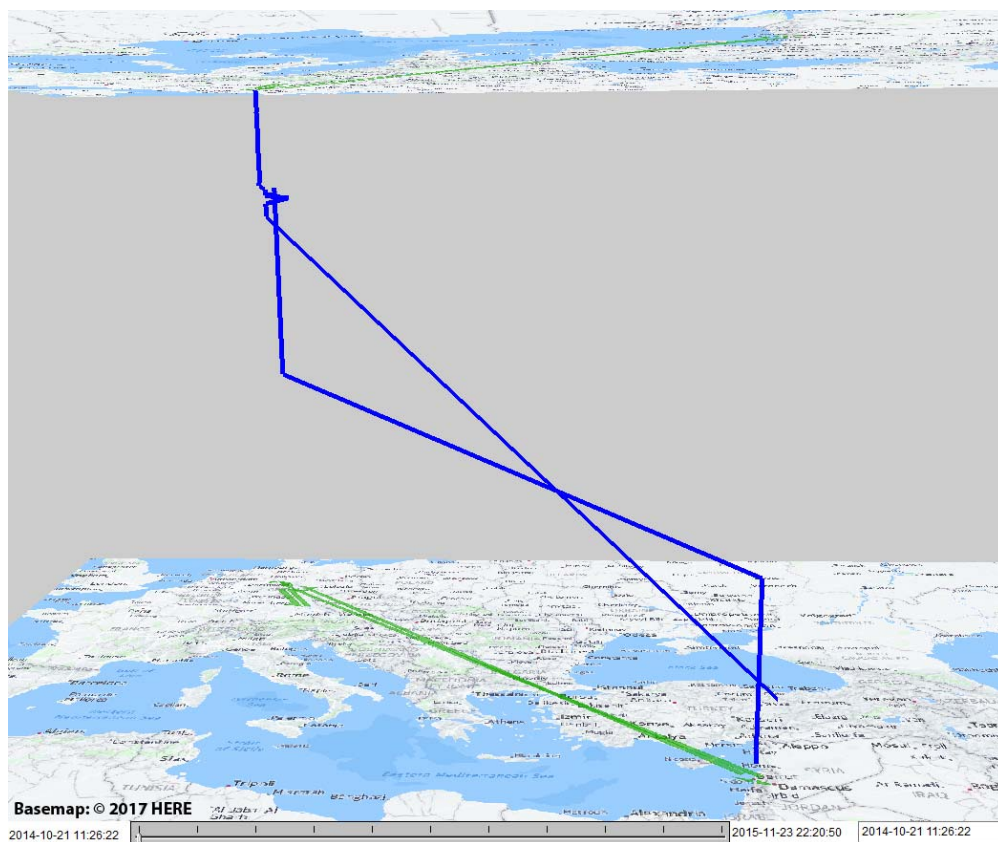

Figure 8. Two likely identified refugee trajectories (blue) and their generalization as routes (green) to Europe visualized in a space-time cube.

The tweet frequency of these two users is shown in Figure 9 over the period of one year. Both patterns show no or only few activities during their potential flight. According to the Twitter time line, a conservative estimation implies that it took the first refugee up to nine months (from December 2014 to August 2015) to reach the destination country and the second refugee two months (from February 2015 to April 2015). Frequent tweets after August 2015 in Germany for both users suggest that Germany was their destination country.
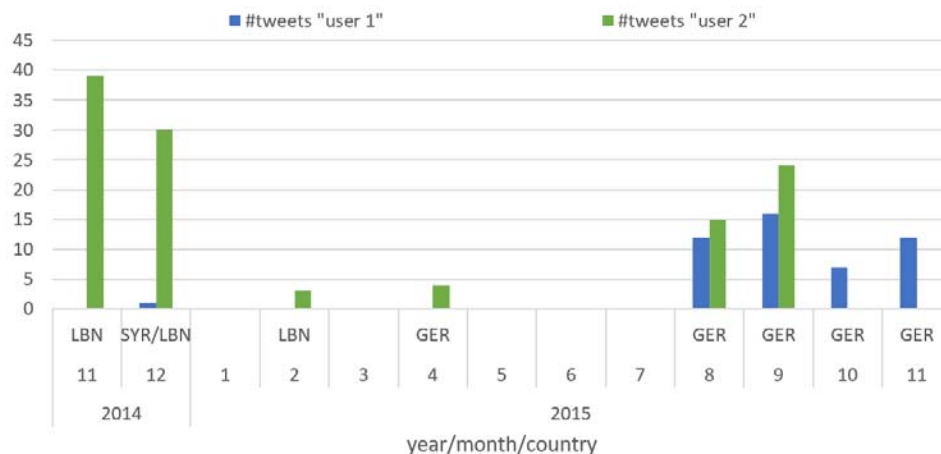

Figure 9. Twitter activity of likely refugees moving from the Middle East to Germany.

5. Local Twitter Activities

In this section, the focus shifts from trajectory analysis to migration related activity detection in selected countries, with general processing steps laid out in Figure 3 (yellow branch). For the process of identifying migration related clusters, tweets were clipped to Germany, Austria, Italy, and Greece before further processing.

5.1. Keyword Lists and Hashtag Extraction

To extract relevant tweets the English keyword list from before (see Section 4.3) was used, however, with terms removed that relate to refugee home countries and political terms. Including the latter two categories would lead to many tweets not directly related to refugees (e.g., a post about the government in Afghanistan). Similar reduced keyword lists were generated for German (e.g., "fluechtlinge" and "asyl"), Italian (e.g., "profugo", "immigrante", and "rifugiato"), and Greek ("prósfygas" and "metanast"). For Austria, Germany, Italy and Greece, the keyword lists in the corresponding languages (German, Italian, and Greek, respectively) were then combined with the reduced English keyword list from before. Next, previously filtered tweets (upper half in Figure 3) were clipped to each of these four countries. For each country, all hashtags were extracted from tweets, and those with matching entries in the corresponding keyword list were retained. After checking and cleaning obtained hashtags, tweets were filtered with hashtags for each country. This process led to the following number of different refugee related hashtags for the four countries (where some of the different hashtags can be quite similar, such as "refugee", "refugee!", or "refugee??"): Germany: 120 in English and 375 in German; Austria: 43 in English and 95 in German; Italy: 76 in English and 219 in Italian; and Greece: 27 in English and 9 in Greek. As an illustration, Figure 10a presents part of the German keyword list used for Austria and Germany, and Figure 10b shows the corresponding word cloud of matching hashtags in analyzed tweets for Austria and Germany.
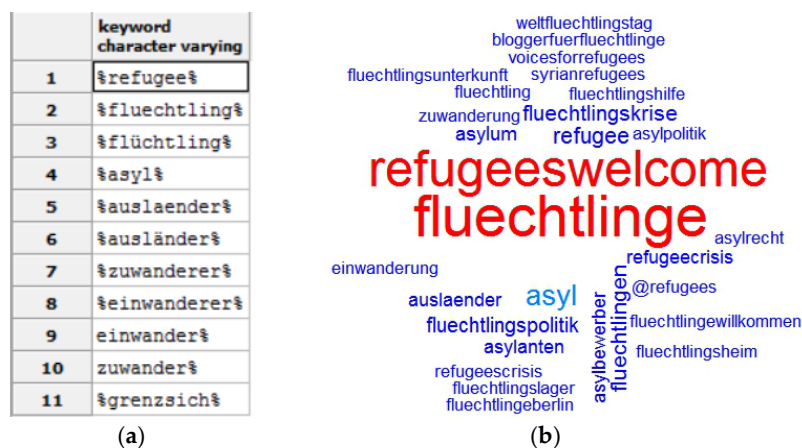
Figure 10. Part of the refugee related keyword list for Austria and Germany (**a**); and corresponding word cloud of hashtags (**b**).

Values in the bar chart in Figure 11a are based on the total number of refugee related tweets posted in each of the four analyzed countries over a period of 13 months (i.e., between November 2014 and November 2015). More specifically, separated by country, each bar indicates the monthly proportion of refugee related tweets within the total set of refugee related tweets. The bar chart suggests a general increase in the number of refugee related tweets in the local population over time. However, a decline in percentage values can be observed in the last two months of the analysis time window following a peak in September 2015. This peak may be the result of refugee related news that triggered additional public response, such as refugees passing through Greece, with highest numbers in September and October 2015 [2]. Similarly, Figure 11b plots for each country the percentage of asylum seekers per month as a proportion of all asylum seekers registered in each country within the same 13-month period, using base data from the UNHCR [46]. Zero values indicate missing data for specific months. The relative frequency of asylum seekers in these four countries continuously rose during that analysis period, which generally matches the perceived pattern in Figure 11a except for those identified peaks in refugee related tweets in September and October 2015.
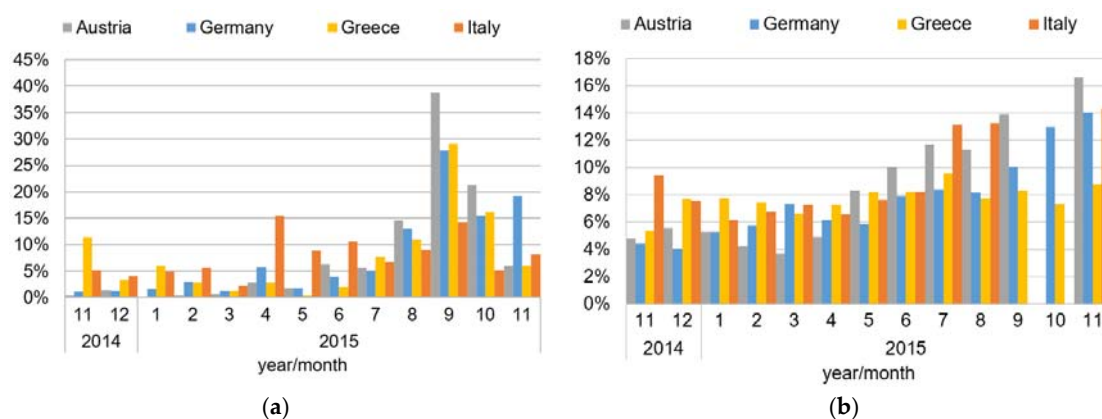

Figure 11. Monthly proportion of refugee related tweets for Austria, Germany, Greece, and Italy out of all refugee related tweets over a 13 month period for each country (**a**); and monthly proportion of asylum seekers for each country out of all registered asylum seekers over the same 13-month period for each country, based on UNHCR data [46] (**b**).

5.2. Cluster Detection

In V-Analytics, points can be defined as events, where each event describes a single object, and not a trajectory. Event clusters can be identified through two methods: "Density-based spatio-temporal event clustering" and "Distance-bounded spatio-temporal event clustering". These methods are described as follows by Cerutti et al. [47], p. 2:

"(1) Density-based clustering detects densely populated regions in space-time with arbitrary shape [. . .]. The number of clusters is not pre-determined and isolated points are optionally discarded as noise, therefore this method is suited for an initial overview and detection of (significant) event candidates."

"(2) Distance-bounded spatio-temporal event clustering [. . .] can be applied to time-dynamic data sets (data streams) and thus can detect emerging spatio-temporal clusters and track their evolution in real-time. This method additionally reconstructs trajectories of clusters, i.e., the evolution of the centre of a cluster's spatial footprint over time."

Method (1) in V-Analytics implements the generic density-based clustering algorithm OPTICS with a suitable distance function [48], and the implementation of Method (2) follows the description given in [18]. Figure 12 shows the results of both methods for Germany and Austria for tweets posted between October 2014 and December 2015. For the OPTICS method a 10,000 m distance threshold (i.e., the maximum allowed distance between neighboring objects), a one-day temporal threshold, and a three-neighbor minimum threshold were chosen, based on visual exploration, analysis scale (city level), and related studies in the literature [47]. For the distance-bounded event clustering the parameter settings are 10,000 m maximum event spatial distance and five as maximum number of events. The randomly colored circles in the space-time cube denote the identified density-based clusters over time (Method (1)). The color scheme is only for visual differentiation and has no additional meaning. The *z*-coordinate expresses the time stamp of a cluster, where the bottom indicates the start date of the observed time range and the top the end. The cluster results suggest that refugee related responses from people picked up from around August 2015, increased for a few months and then slightly declined towards the end of the year, generally supporting the trend of the proportions of refugee related tweets observed in Figure 11a. The red circles that are projected onto the maps on the bottom and top of the space-time cube and onto the inset map denote the distance-bounded event clusters (Method (2)). They indicate areas of increased Twitter activities related to refugee topics. Large clusters were identified for Berlin, Hanover, Leipzig, Essen, Cologne and Munich (Germany), and Vienna (Austria). The abundance of event clusters identified for Germany and Austria demonstrates that in this study space-centered tasks result in more informative data points related to refugee movements than trajectory analysis, however, at the cost of losing travel flow information.
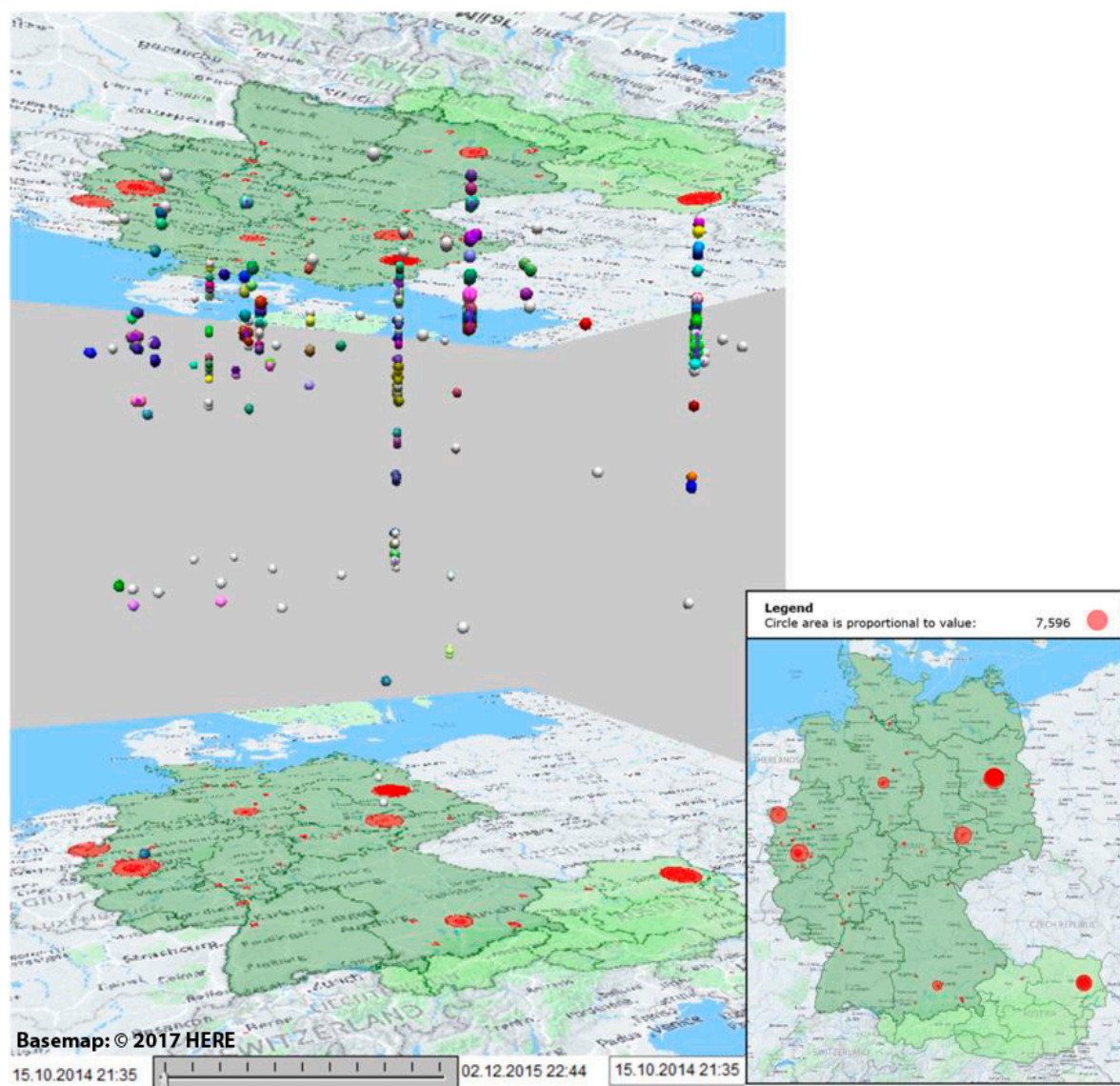
Figure 12. Density-based and distance-bounded spatio-temporal event clustering for Austria and Germany visualized in a space-time cube.

5.3. *Local Analysis of Activity Patterns*

Presence of tweets with specific thematic attributes, optionally combined with spatio-temporal information, can reveal patterns of refugee related activities at the local level, which will be examined for Austria, Germany, and Greece in detail.

5.3.1. Austria

The proportion of refugee related tweets in the nine Austrian provinces (computed for October 2014 to December 2015) ranges between about 0.003 and 0.088 percent. The map in Figure 13 suggests that the proportions are higher in the eastern provinces, with the highest rate observed in Burgenland. The latter can be likely explained by the fact that many refugees entered the EU via Greece, continued their journey through several Balkan countries and reached Austria via Hungary, which borders Burgenland to the east. Mid 2015, thousands of refugees either walked or were transported by bus from refugee camps in Hungary to the Austrian border [49], which was a widely discussed topic in the news and social media in Austria around that time. When Hungary closed the border

with Croatia to migrants around mid-October 2015, thousands of migrants entered Slovenia instead, many of whom continued their journey to Austria from the South, mostly into Styria [50]. This change in refugee routes through Slovenia falls, however only into the last two months of our analysis time frame, and thus is expected to have a smaller effect on refugee related tweet percentages than arrivals through Hungary. Results of a linear regression that predicts the percentage of refugee related tweets in Austrian provinces (Table 1) finds that the distance between a province centroid and the Hungarian border is a strong predictor of the percentage of refugee related tweets, resulting in an adjusted R^2 of 0.63. Using the nearest distance to either the Hungarian or Slovenian border results in a lower adjusted R^2 of 0.52, suggesting that for the analysis time frame refugee flows through Hungary had a more significant effect on social media activities in Austria than refugee flows through Slovenia later on.
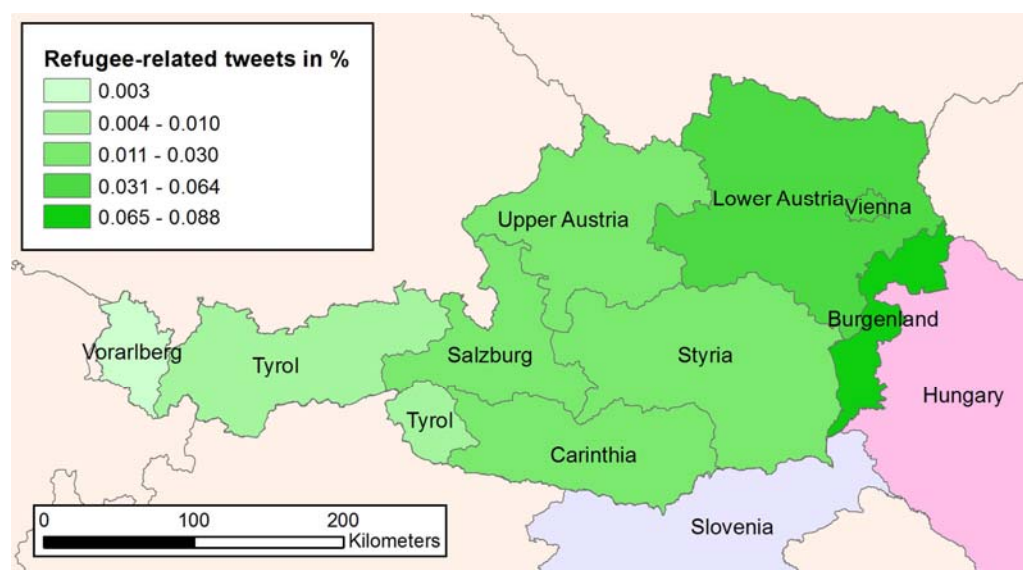

Figure 13. Percentage of refugee related tweets in Austrian provinces.

Table 1. Estimation results for percentage of refugee related tweets in Austrian provinces.

	Coefficient	Std. Err.	t	Sig.
Constant	0.630	0.009	6.937	0.000 **
Distance to Hungary (in 1000s of km)	−0.149	0.039	−3.848	0.006 **
N	9			
Adjusted R^2	0.633			

Note: ** $p < 0.01$, * $p < 0.05$.

5.3.2. Germany

The percentage of refugee related tweets for the 16 federal states in Germany is mapped in Figure 14. The highest value is found for Brandenburg (0.244%), followed by Berlin (0.111%). The lowest values are found for Hamburg (0.025%) and Mecklenburg-West Pomerania (0.027%).

Brandenburg is an outlier with a refugee related tweet rate more than double that of the second highest rate for Berlin. Unemployment is high in some cities in the east, increasing the aversion of some local residents against more immigration in fear of loss of jobs to migrants. An example is Frankfurt an der Oder, located in east Brandenburg and bordering Poland. This city has an unemployment rate of 10.0%, compared to about 5.9% nationwide. Although foreigners make up only about one percent of the population, there is sentiment within the population that this rate is too high [51]. Such concerns may have contributed to the observed higher refugee related tweet rate in Brandenburg than in other federal states.
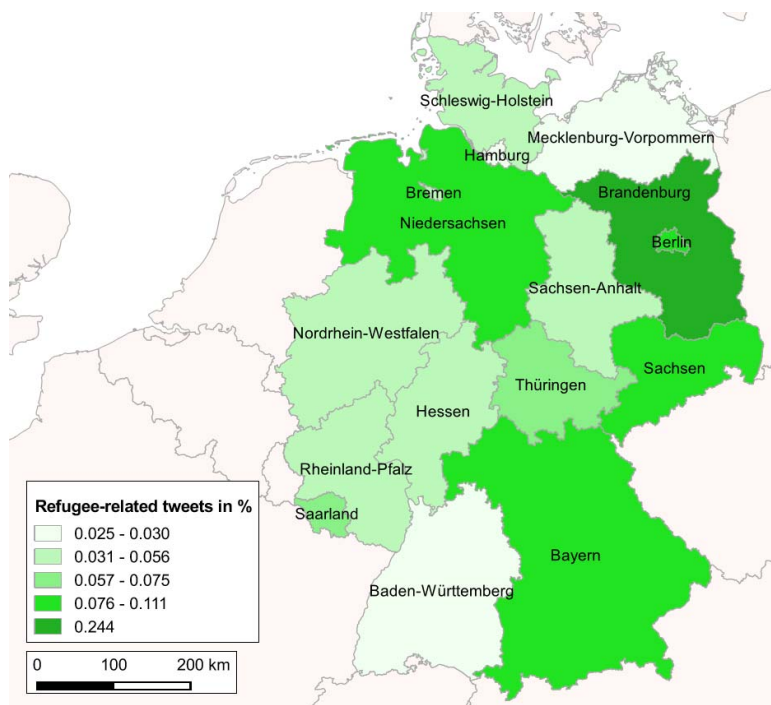
Figure 14. Percentage of refugee related tweets in federal states in Germany.

5.3.3. Greece

Figure 15 shows the proportion of refugee related tweets in decentralized public authorities (provinces) of Greece throughout the analysis year. These tweets are primarily posted by the local population and dedicated support teams. The highest proportion value can be observed on the eastern-most islands, such as Lesbos or Chios, where most refugees from the Middle East arrived after their hazardous trips from Turkey [52].
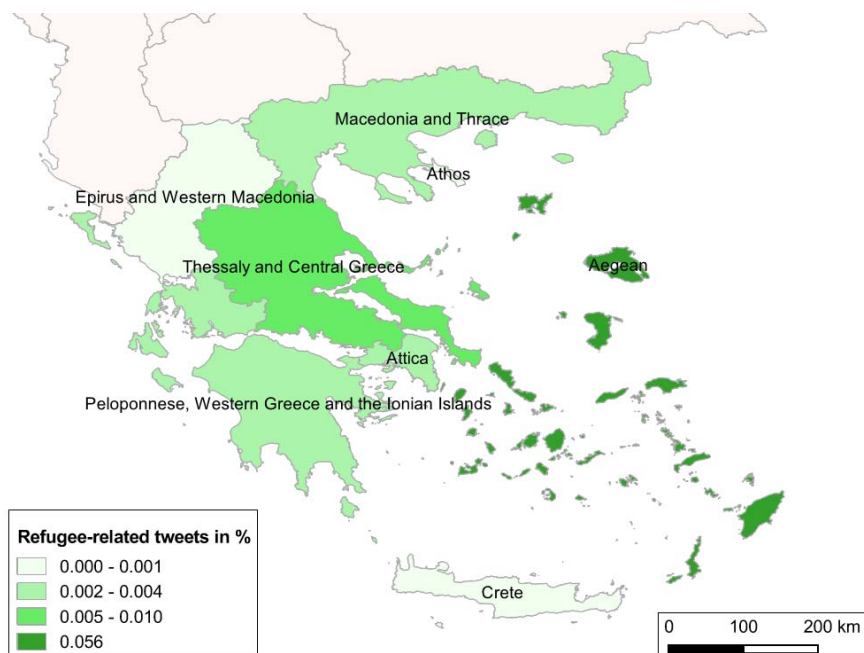

Figure 15. Percentage of refugee related tweets in decentralized public authorities for Greece.

Figure 16 shows the clusters identified through the distance-bounded event clustering method in V-Analytics. The clusters are primarily found in areas around Corfu, Athens and Lesbos, which are generally smaller cities except for Greece's capital, Athens. In fact, only two cities with identified clusters nearby are within the 10 largest cities in Greece. These are Athens (rank: 1, population: 3.2 million) and Volos (rank: 6; population: 130,000). This means that tweets do indicate locations of refugee related activities, even if they have a small number of residents. This is supported by the fact that the cluster locations match well with refugee routes that run within the Aegean, Ionian, and Adriatic Sea (Figure 17). For example, refugees are known to board ships illegally in Corfu to reach Central Europe [53,54]. Lesbos is holding major refugee camps, e.g., the Moria Refugee Camp, and Athens hosts several institutions and community centers related to refugee management and support.
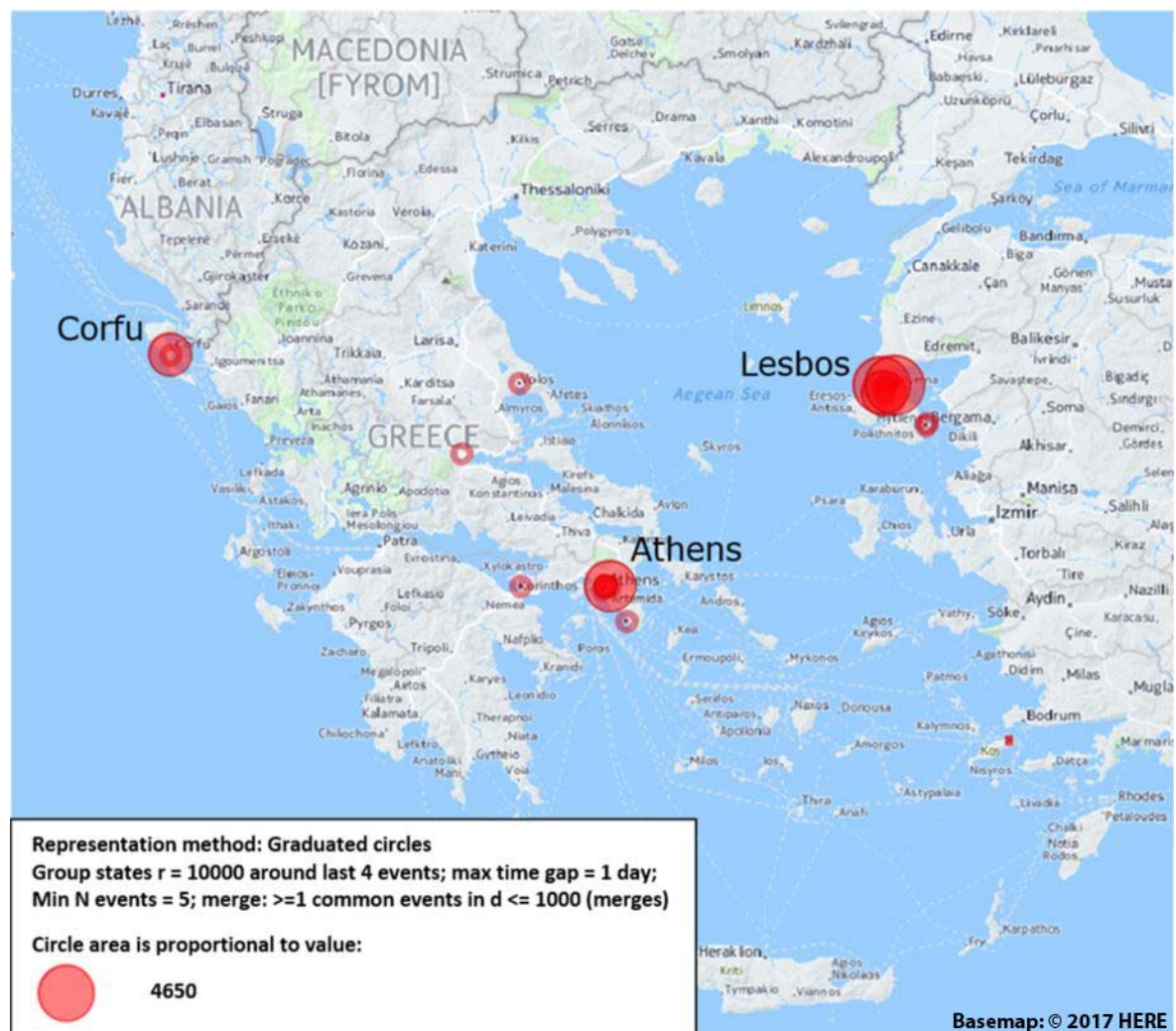

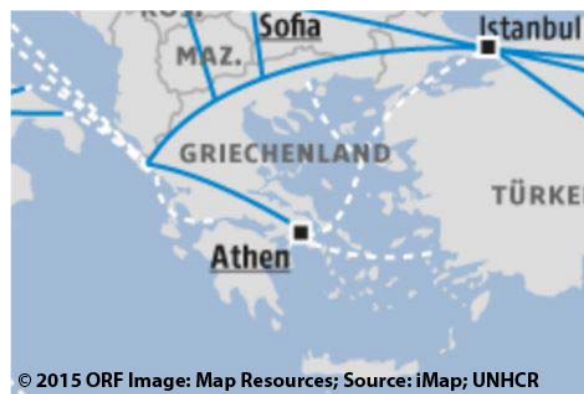**Figure 16.** Result of distance-bounded event clustering for Greece.

Figure 17. Primary refugee routes along the Mediterranean Sea, extract for Greece (adapted from [55]).

6. Summary and Discussion

This paper explored different methods to identify refugee migration patterns from the Middle East and Africa to Europe based on geo-tagged tweets. To the best of our knowledge, this is the first study to use tweets for this purpose. Using the V-Analytics software tools in combination with SQL queries for data preparation and filtering, and motivated by the possibility to view crowd-sourced space- and time-referenced data in a dual way [19,20], the study presented two objectives of refugee pattern analysis: (1) trajectory filtering and aggregation; and (2) keyword-based event detection.

Regarding the first objective, the study demonstrated that the exploration of trajectory filter settings in combination with the manual classification of tweet messages based on tweet content was necessary to obtain even a small set of trajectories that most likely originate from refugees. For this purpose, the graphical user interface and immediate visual output in V-Analytics was helpful to improve filter settings in an iterative approach. Although several parameter settings were tested and presented for SQL-based trajectory filtering (e.g., maximum distance between two tweets along a trajectory) and trajectory aggregation in V-Analytics (e.g., minimum number of trajectories per arrow), the filter and aggregation settings are not necessarily transferable to other case studies related to refugee movement analysis. This is because optimum parameter settings will depend on varying factors, such as data abundance, geographic scale, topic, or transportation mode [42]. As demonstrated in this paper, working with scarce spatio-temporal data requires iterative exploration and manual review of available information to obtain tangible results. Manual content search was also found to be a necessary step in other Twitter based studies to obtain relevant data [45]. The challenges faced in the presented study demonstrated that big data from social media platforms alone does not necessarily contain a high information value in its raw state for a given task, and requires multi-faceted methods for relevant information extraction. This study contributes therefore to the general topic of "Big Noise", which has been described in other related studies of movement analysis (not in the context of migration, though). One study, for example, cross-compared three estimates for human movements from home locations to shopping centers that were derived from a major telephone provider, a commercial consumer survey, and geo-tagged tweets [56]. Results showed that tweets produce the fewest flow data and diverged from other sources, illustrating problems of data scarcity and accuracy of geo-tagged tweets for selected movement related tasks. In the current study the low detection rate of trajectories from the Middle East and Africa to Europe has also to do with the generally low Twitter penetration rate of countries in the Middle East, such as Syria, Afghanistan, and Iraq, as well as Northern African countries [17], possibly due to limited data and Wi-Fi access on mobile devices in certain regions in the world [57]. Whatever the Twitter penetration rate is, the scarcity of geo-tagged tweets (only about 1% among all tweets) poses an additional challenge for conducting spatial event and movement analysis. To compensate for the scarcity of geo-tagged tweets, several studies have explored a variety of methods to geo-locate tweets through other sources of information

in the tweet post or in the user profile, such as tweet content and social network structure [58,59]. Although these position estimation methods are consistently improving, they add a level of positional uncertainty to any subsequent analysis. The efficacy of these methods could also be explored in the context of migration pattern analysis with the aim to increase the number of relevant trajectories. This would, however, exceed the scope of this paper, and can therefore be considered a part of future work. Another aspect to explore in future work is the use of mobile phone tracking data to obtain a more complete picture of refugee movements, possibly in combination with tweets. Phone based positioning does not depend on a specific application platform, such as Twitter or Foursquare/Swarm. However, obtaining a cohesive set of spatio-temporal locations from phone records across multiple countries appears to be challenging, given that most studies that analyze mobility from phone records are geographically limited to the city level [60,61] or country level [62], with only a few exceptions that extend beyond national or continental boundaries [63].

The second objective of this study, i.e., hashtag based cluster detection, yielded more informative information about migration routes than the trajectory extraction approach. Parameters were iteratively chosen, using the geographic analysis scale and parameter settings found in related studies as a starting point. The obtained clusters of refugee related tweets occur primarily at refugee camps or major transfer points along common refugee flight routes, confirming hot spots reported in various online media. Interestingly, the clusters lie mostly outside large cities, showing that population density does not strongly bias results of this analysis method. Density-based and distance-bounded spatio-temporal event clustering methods result also in information about the frequency of refugee clusters at certain locations and therefore allow to draw a general picture of the development of topical trends over time in a geographic region of interest. Comparison between the monthly proportion of hashtags mentions relating to refugees and the number of asylum seekers in selected countries, as provided by the UNHCR, shows a good match in trends. Since authoritative data about refugees are often comprised and presented only in aggregated form at the state or federal level, Twitter based event clustering can complement this information. Local analysis revealed a spike in migration related tweets for parts of Germany, and also demonstrated a decline in refugee related activities from east to west in Austria. The latter can be well explained by known migration routes to Europe. Sentiment analysis of refugee related tweets [64,65] in identified clusters might reveal changing attitudes of the local population towards refugees over time. Using full text analysis, e.g., LDA, could be used for a refined classification of extracted topics of tweets along suspected refugee routes [66].

Author Contributions: H. H. and G. P. developed the research topic. S. C. downloaded and pre-processed the Twitter data. F. H. designed the analysis work flow, analyzed the data, visualized the results, and wrote the first draft. H. H. edited later versions of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saarinen, V.; Ojala, J. The Flow towards Europe. 2017. Available online: http://www.lucify.com/the-flow-towards-europe/ (accessed on 5 July 2017).
2. UNHCR. Global Trends—Forced Displacement in 2015. 2015. Available online: http://www.unhcr.org/576408cd7.pdf (accessed on 10 August 2017).
3. Robinson, D. How the EU Plans to Overhaul "Dublin Regulation" on Asylum Claims. 2016. Available online: https://www.ft.com/content/d08dc262-bed1-11e5-9fdb-87b8d15baec2 (accessed on 29 July 2017).
4. The Telegraph. Refugee Crisis: Many Migrants Falsely Claim to be Syrians, Germany Says as EU Tries to Ease Tensions. 2015. Available online: http://www.telegraph.co.uk/news/worldnews/europe/germany/11891219/Refugee-crisis-Many-migrants-falsely-claim-to-be-Syrians-Germany-says-as-EU-tries-to-ease-tensions.html (accessed on 29 July 2017).
5. Wood, D. *The Power of Maps*; The Guilford Press: New York, NY, USA, 1992.

6.   Monmonier, M. *How to Lie with Maps*; The University of Chicago Press: Chicago, IL, USA, 1991.

7.   Quam, L.O. The Use of Maps in Propaganda. *J. Geogr.* **1943**, *42*, 21–32. [CrossRef]

8.   Li, L.; Goodchild, M.F.; Xu, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 61–77. [CrossRef]

9.   Bittner, C. Diversity in volunteered geographic information: Comparing OpenStreetMap and Wikimapia in Jerusalem. *GeoJournal* **2016**. [CrossRef]

10.   Lotan, G.; Graeff, E.; Ananny, M.; Gaffney, D.; Pearce, I.; Boyd, D. The Revolutions Were Tweeted: Information Flows during the 2011 Tunisian and Egyptian Revolutions. *Int. J. Commun.* **2011**, *5*, 1375–1406.

11.   Pei, S.; Muchnik, L.; Andrade José, S.J.; Zheng, Z.; Makse, H.A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **2014**, *4*, 5547. [CrossRef] [PubMed]

12.   Graham, M.; Hale, S.A.; Gaffney, D. Where in the World Are You? Geolocation and Language Identification in Twitter. *Prof. Geogr.* **2014**, *66*, 568–578. [CrossRef]

13.   Azmandian, M.; Singh, K.; Gelsey, B.; Chang, Y.-H.; Maheswaran, R. Following Human Mobility Using Tweets. In *Agents and Data Mining Interaction (LNCS Volume 7607)*; Cao, L., Zeng, Y., Symeonidis, A.L., Gorodetsky, V.I., Yu, P.S., Singh, M.P., Eds.; Springer: Berlin, Germay, 2013; pp. 139–149. [CrossRef]

14.   Krumm, J.; Caruana, R.; Counts, S. Learning Likely Locations. In *User Modeling, Adaptation, and Personalization—Proceedings of UMAP 2013 (LNCS 7899)*; Carberry, S., Weibelzahl, S., Micarelli, A., Semeraro, G., Eds.; Springer: Berlin, Germay, 2013; pp. 64–76. [CrossRef]

15.   Valle, D.; Cvetojevic, S.; Robertson, E.P.; Reichert, B.E.; Hochmair, H.H.; Fletcher, R.J. Individual Movement Strategies Revealed through Novel Clustering of Emergent Movement Patterns. *Sci. Rep.* **2017**, *7*, 44052. [CrossRef] [PubMed]

16.   Lenormand, M.; Gonçalves, B.; Tugores, A.; Ramasco, J.J. Human diffusion and city influence. *J. R. Soc. Interface* **2015**, *12*, 20150473. [CrossRef] [PubMed]

17.   Hawelka, B.; Sitko, I.; Beinat, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. Geo-located Twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 260–271. [CrossRef] [PubMed]

18.   Andrienko, N.; Andrienko, G.; Fuchs, G.; Rinzivillo, S.; Betz, H.-D. Detection, Tracking, and Visualization of Spatial Event Clusters for Real Time Monitoring. In *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA)*; IEEE: Paris, France, 2015. [CrossRef]

19.   Andrienko, G.; Andrienko, N.; Bak, P.; Kisilevich, S.; Keim, D. Analysis of Community-Contributed Space- and Time-Referenced Data (Example of Flickr and Panoramio Photos). In Proceedings of the 2009 IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, NJ, USA, 12–13 October 2009; pp. 213–214. [CrossRef]

20.   Andrienko, G.; Andrienko, N.; Bak, P.; Keim, D.; Wrobel, S. *Visual Analytics of Movement*; Springer: Heidelberg, Germany, 2013.

21.   von Landesberger, T.; Bremm, S.; Schreck, T.; Fellner, D.W. Feature-based automatic identification of interesting data segments in group movement data. *Inf. Vis.* **2013**, *13*, 190–212. [CrossRef]

22.   Romanillos, G.; Zaltz Austwick, M.; Ettema, D.; De Kruijf, J. Big Data and Cycling. *Transp. Rev.* **2016**, *36*, 114–133. [CrossRef]

23.   Alivand, M.; Hochmair, H.H.; Srinivasan, S. Analyzing how travelers choose scenic routes using route choice models. *Comput. Environ. Urban Syst.* **2015**, *50*, 41–52. [CrossRef]

24.   Beiró, M.G.; Panisson, A.; Tizzoni, M.; Cattuto, C. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Sci.* **2016**, *5*, 30. [CrossRef]

25.   Sun, Y.; Fan, H.; Bakillah, M.; Zipf, A. Road-based travel recommendation using geo-tagged images. *Comput. Environ. Urban Syst.* **2015**, *53*, 110–122. [CrossRef]

26.   Rösler, R.; Liebig, T. Using Data from Location Based Social Networks for Urban Activity Clustering. In *Geographic Information Science at the Heart of Europe*; Vandenbroucke, D., Bucher, B., Crompvoets, J., Eds.; Springer International Publishing: Cham, Switzerland, 2013; pp. 55–72. [CrossRef]

27.   Lenormand, M.; Tugores, A.; Colet, P.; Ramasco, J.J. Tweets on the road. *PLoS ONE* **2014**, *9*, e105407. [CrossRef] [PubMed]

28.   Steiger, E.; de Albuquerque, J.P.; Zipf, A. An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter Data. *Trans. GIS* **2015**, *19*, 809–834. [CrossRef]

29. Senaratne, H.; Broering, A.; Schreck, T.; Lehle, D. Moving on Twitter: Using Episodic Hotspot and Drift Analysis to Detect and Characterise Spatial Trajectories. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*; ACM Press: New York, NY, USA, 2014; pp. 23–30.

30. Shelton, T.; Poorthuis, A.; Graham, M.; Zook, M. Geoforum Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. *Geoforum* **2014**, *52*, 167–179. [CrossRef]

31. Crooks, A.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. #Earthquake: Twitter as a Distributed Sensor System. *Trans. GIS* **2013**, *17*, 124–147. [CrossRef]

32. Cassa, C.A.; Chunara, R.; Mandl, K.; Brownstein, J.S. Twitter as a Sentinel in Emergency Situations: Lessons from the Boston Marathon Explosions. *PLOS Curr. Disasters* **2013**, *2*, 1–11. [CrossRef] [PubMed]

33. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proceedings of the 19th International Conference on World Wide Web*; ACM: New York, NY, USA, 2010; pp. 851–860. [CrossRef]

34. Zagheni, E.; Garimella, V.R.K.; Weber, I.; State, B. Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*; ACM: New York, NY, USA, 2014; pp. 439–444. [CrossRef]

35. Rüegger, S.; Bohnet, H. The Ethnicity of Refugees (ER): A new dataset for understanding flight patterns. *Confl. Manag. Peace Sci.* **2015**. [CrossRef]

36. Iqbal, Z. The Geo-Politics of Forced Migration in Africa, 1992—2001. *Confl. Manag. Peace Sci.* **2007**, *24*, 105–119. [CrossRef]

37. Rettberg, J.W.; Gajjala, R. Terrorists or cowards: Negative portrayals of male Syrian refugees in social media. *Fem. Media Stud.* **2016**, *16*, 178–181. [CrossRef]

38. Darwish, K.; Magdy, W. Attitudes towards Refugees in Light of the Paris Attacks. 2015. Available online: https://arxiv.org/abs/1512.04310 (accessed on 20 July 2017).

39. Roesslein, J. Tweepy Documentation [Internet]. 2009. Available online: http://docs.tweepy.org/en/v3.5.0/ (accessed on 20 June 2017).

40. Uddin, M.M.; Imran, M.; Sajjad, H. Understanding Types of Users on Twitter. *arXiv Prepr.* 2014. Available online: https://arxiv.org/abs/1406.1335 (accessed on 11 July 2017).

41. Zhang, C.M.; Paxson, V. Detecting and Analyzing Automated Activity on Twitter. In *Passive and Active Measurement, PAM 2011*; Spring, N., Riley, G., Eds.; Springer: Berlin, Germany, 2011; pp. 102–111. [CrossRef]

42. Andrienko, N.; Andrienko, G. Spatial generalisation and aggregation of massive movement data. *IEEE Trans. Vis. Comput. Graph.* **2011**, *17*, 205–219. [CrossRef] [PubMed]

43. Andrienko, G.; Andrienko, N.; Wrobel, S. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explor. Newsl.* **2007**, *9*, 38. [CrossRef]

44. Chong, M. Sentiment analysis and topic extraction of the twitter network of #prayforparis. *Proc. Assoc. Inf. Sci. Technol.* **2016**, *53*, 1–4. [CrossRef]

45. Guzman, E.; Alkadhi, R.; Seyff, N. A Needle in a Haystack: What Do Twitter Users Say about Software? In Proceedings of the 2016 IEEE 24th International Requirements Engineering Conference (RE), Beijing, China, 12–16 September 2016; pp. 96–105. [CrossRef]

46. UNHCR. *UNHCR Population Statistics—Data—Time Series.* 2017. Available online: http://popstats.unhcr.org/en/time_series (accessed on 7 August 2017).

47. Cerutti, V.; Fuchs, G.; Andrienko, G.; Andrienko, N.; Ostermann, F. *Identification of Disaster-Affected Areas Using Exploratory Visual Analysis of Georeferenced Tweets: Application to a Flood Event*; Association of Geographic Information Laboratories in Europe: Helsinki, Finland, 2016; p. 5.

48. Andrienko, G.; Andrienko, N.; Rinzivillo, S.; Nanni, M.; Pedreschi, D.; Giannotti, F. Interactive visual clustering of large collections of trajectories. In *Visual Analytics Science and Technology (VAST)*; IEEE: Atlantic City, NJ, USA, 2009; pp. 3–10. [CrossRef]

49. The Guardian. Hungary to Take Thousands of Refugees to Austrian Border by Bus. 2015. Available online: https://www.theguardian.com/world/2015/sep/04/hundreds-refugees-march-austria-budapest-hungary-syrians (accessed on 31 July 2017).

50. BBC. Migrant crisis: Thousands Enter Slovenia after Hungary Closes Border. 2015. Available online: http://www.bbc.com/news/world-europe-34564830 (accessed on 30 July 2017).

51. The Local. Few Foreigners in Eastern Germany but Xenophobia Is Rife. 2017. Available online: https://www.thelocal.de/20170326/few-foreigners-in-eastern-germany-but-xenophobia-is-rife (accessed on 29 July 2017).

52. Leadbeater, C. Which Greek Islands are Affected by the Refugee Crisis? 2016. Available online: http://www.telegraph.co.uk/travel/destinations/europe/greece/articles/greek-islands-affected-by-refugee-crisis/ (accessed on 31 July 2017).

53. Associated Newspapers Ltd. Italian Coastguard Seizes Cargo Ship Carrying 600 Illegal Migrants after the Crew Programmed the Vessel to Crash into Coast before Fleeing. 2014. Available online: http://www.dailymail.co.uk/news/article-2891118/Ship-coast-Corfu-carrying-700-passengers-issues-SOS-armed-men-board.html (accessed on 22 April 2017).

54. Telegraph Media Group Ltd. Mysterious Migrant "Ghost Ship" Arrives in Italy. 2014. Available online: http://www.telegraph.co.uk/news/worldnews/europe/italy/11318586/Mysterious-migrant-ghost-ship-arrives-in-Italy.html (accessed on 22 April 2017).

55. ORF. Wichtige Flüchtlingsrouten. 2016. Available online: http://orf.at/stories/2307356/2307294/ (accessed on 5 August 2017).

56. Lovelace, R.; Birkin, M.; Cross, P.; Clarke, M. From Big Noise to Big Data: Toward the Verification of Large Data sets for Understanding Regional Retail Flows. *Geogr. Anal.* **2016**, *48*, 59–81. [CrossRef]

57. Cvetojevic, S.; Juhász, L.; Hochmair, H.H. Positional Accuracy of Twitter and Instagram Images in Urban Environments. *GI_Forum* **2016**, *1*, 191–203. [CrossRef]

58. Cheng, Z.; Caverlee, J.; Lee, K. You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, 26–30 October 2010; pp. 759–768. [CrossRef]

59. Kotzias, D.; Lappas, T.; Gunopulos, D. Addressing the Sparsity of Location Information on Twitter. In Proceedings of the Workshop of the EDBT/ICDT 2014 Joint Conference, Athens, Greece, 28 March 2014.

60. Sagl, G.; Delmelle, E.; Delmelle, E. Mapping collective human activity in an urban environment based on mobile phone data. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 272–285. [CrossRef]

61. Lenormand, M.; Picornell, M.; Cantú-Ros, O.G.; Tugores, A.; Louail, T.; Herranz, R.; Barthelemy, M.; Frías-Martínez, E.; Ramasco, J.J. Cross-Checking Different Sources of Mobility Information. *PLoS ONE* **2014**, *9*, e105184. [CrossRef] [PubMed]

62. Lu, X.; Wrathall, D.J.; Sundsøy, P.R.; Nadiruzzaman, M.; Wetter, E.; Iqbal, A.; Qureshi, T.; Tatem, A.; Canright, G.; Engø-Monsen, K.; et al. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Glob. Environ. Chang.* **2016**, *38*, 1–7. [CrossRef]

63. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [CrossRef] [PubMed]

64. Dwibhasi, S.; Jami, D.; Lanka, S. Analyzing and Visualizing the Sentiments of Ebola Outbreak Via Tweets. In Proceedings of the SAS Global Forum, Dallas, TX, USA, 26–29 April 2015; pp. 1–12.

65. Mitchell, L.; Frank, M.R.; Harris, K.D.; Dodds, P.S.; Danforth, C.M. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE* **2013**, *8*, e64417. [CrossRef] [PubMed]

66. Steiger, E.; Resch, B.; Zipf, A. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1694–1716. [CrossRef]