ISSN 1870-4069

Regular Activity Patterns in Spatio-Temporal Events Databases: Multi-Scale Extraction of Geolocated Tweets

Pablo López-Ramírez, Alejandro Molina-Villegas, Oscar Sánchez-Siordia,
Mario Chirinos-Colunga, Gandhi Hernández-Chan

Conacyt - Centro de Investigación en Ciencias de Información Geoespacial,
Mexico

Abstract. This paper proposes a new technique for the extraction of regular activity patterns at different scales (resolution levels), mined from the microblogging platform Twitter. The approach is based on the recursive application of the DBSCAN clustering algorithm to the geolocated Twitter feed. The proposed technique includes a novel way to obtain 'averaged' regular activity zones based on the rasterization and aggregation of the Concave Hull of the clusters identified at each resolution level. This technique uses only the spatio-temporal characteristics of the geolocated Twitter feed and does not depend on the data content; therefore it can be extended to work with different spatio-temporal event sources such as mobile telephone records. An experiment was carried out to demonstrate the effectiveness of our technique in the extraction of known activity patterns in the Mexico City Metropolitan Area.

Palabras clave: social media, urban activity, geographic data mining.

1 Introduction

Spatio-Temporal analysis is a rapidly growing field within Geographical Information Science (GIS). The rate of increase in the amount of information gathered every day, the pervasiveness of Global Positioning System (GPS) enabled sensors, mobile phones, social networks and the Internet of Things (IoT), demand for robust and efficient analysis techniques that can help us find meaningful insights from large spatio-temporal databases. Within these new sources of information, the digital breadcrumbs left behind by social media users, have proven to be a valuable resource. They allow us to examine different aspects of crowd behavior, for example, the role of this new media in the Arab Spring [16], or the way people react to hazardous events in general [24] or to more specific occurrences, like terrorist attacks [23] or earthquakes [5].

In the GIS field, one of the main lines of research has been the detection of events [3] or the characterization of zones through social media activity [11], [20]. In both cases, extraction and characterization of regular activity patterns is very important. With this in mind, in this paper we propose a technique for the extraction of regular patterns of activity that relies only on the spatio-temporal features of the Tweeter feed. The purpose of this is, on the one hand, improve on the current available techniques [21], [11] and, on the other hand, to be as less dependent as possible from the nature of the Twitter feed.

The proposed technique is based on the observation that the regular activity patterns exhibit a wide range of scales [2], and that the current methods for determining this activity from Twitter messages do not consider this. The proposed approach is based on the recursive application of a clustering algorithm to extract patterns of activity across several scales or resolution levels. This approach demands the development of a novel way for 'averaging' the spatial patterns (clusters) extracted from the data.

The rest of the paper is organized as follows: in Section §2 we will establish the basic concepts and perform a general review of the available literature. In Section §3 we will explain in detail the proposed technique. Section §4 presents an experiment extracting the regular activity patterns at several scales from the geolocated Twitter feed in Central Mexico. Finally, Section §5 concludes this document.

2 Spatio-Temporal Events and Crowd Activity Detection

In the context of spatio-temporal events, as described by Kisilevich et. al [18], we refer as Crowd Activity to the collective aggregated patterns observed in some spatio-temporal events datasets, specially in data describing some aspect of the behavior of human populations. Although not formally defined, this concept underpins most of the work that we are going to review in the rest of this section.

Moving on to the subject of using Twitter as a source of geographical insights, there is a substantial body of work on techniques for event [1] detection using Twitter. Atefeh and Khreich [3] present a survey of such techniques. It is interesting to note that most of the work done on the subject has been particularly focused on extracting information from the content of the messages; this is natural since it is to be expected that the written messages contain valuable information that can be analyzed to extract meaningful insights. On the other hand, extracting meaningful information from the messages on Twitter can be a daunting task, not only because of the complications associated with analyzing natural language, but also because the Twitter feed is known to be polluted with meaningless messages, rumors, bots, and other kinds of spurious content [14],[20].

There is also an important amount of work done on using geolocated tweets to extract information about the geographic environment of the users. Gabrielli, et. al. [13] propose a framework for mining anomalous mobility patterns using geolocated tweets, they track the movement of users and semantically enrich their trajectories with information about the users and the kinds of activity present at the different locations each user has visited. Kim, et. al. [17] provide

[1] In this context, event refers to 'real world occurrences that unfold over space and time' [3], which is different to the use of the term on spatio-temporal databases.

a methodology for finding user clusters on nearby locations; they use both the geographic location and the content of the messages for detecting spontaneous events associated with topics of interest. In the same line, Boettcher and Lee [4] present a refinement of local spontaneous event detection that uses historic data to assess the regular topics related to a specific area.

All of the works cited above have in common the use of both the geographic location and the semantic content of the messages. However, there is also some work done on the extraction of meaningful patterns using only the spatio-temporal properties of the data and disregarding the semantic dimension completely. In this latter category, Frias-Martinez, et. al. ([10] and [11]) propose a technique for detecting land use by analysing geolocated tweets. This work is particularly relevant because it involves the extraction of regular patterns of activity from the geolocated Twiter feed and will be discussed in Section 3.

On the topic of unusual activity detection, Lee, et. al. [21], Fujisaka, et. al. [12] and Lee and Sumiya [20], propose successive refinements of a technique for detecting unusually crowded places extracting the regular pattern of activity by performing K-Means clustering over the geolocated tweets and then characterizing each cluster by the number of users, the amount of messages and a measure of the mobility of users in each cluster. Next, the unusual activity is detected by comparing the regular pattern with the characteristics of a specific moment.

From this brief review we can infer some generalities involved in the construction of regular patterns of activity from the geolocated Twitter feed:

- Time is segmented in intervals and the definition of this intervals is arbitrary. In [11], each day is divided in 20 minutes intervals, while in [21], each day is segmented in four six hours intervals.
- The geographic space is partitioned in a flat cluster hierarchy. Frias-Martinez, et. al use a Self Organizing Map [19] to obtain a tessellation of the study area, while Lee, et. al. use K-Means to obtain a similar tessellation.

This work will be focused on the second general characteristic: the way in which the space is partitioned to obtain regular activity zones. In both cases, the partition algorithm returns a flat hierarchy of zones which is a Voronoi tessellation around the cluster centroids identified. This partition reflects the differences in event (point) density across the whole space but, since it is flat (it has a single hierarchic level) it cannot represent the structures found at different scales, this means that such partitions mix the whole range of scales of the underlying processes into a single tessellation.

However, when addressing the regular activity patterns from a geographic perspective, the issue of scale is evident: the underlying processes that generate the observed spatio-temporal distribution of events are organized as a hierarchy of scales. In the case of geolocated tweets, the underlying process is the daily pattern of activity, whether in a whole region (as in [21]) or in a urban zone (in [11]). These patterns are closely related to the general fabric of the city or the region and those have been shown to exhibit different properties when analysed at different scales [2]. This suggests that the techniques for defining the regular activity patterns can be improved by explicitly incorporating the concept of scale.

3 Multi-Scale Regular Activity Extraction

The main idea behind the technique proposed in this paper, is that the regular patterns of activity exhibit a range of scales, and that these scales cannot be represented by a flat tessellation. To overcome this limitation, we propose the use of a recursive algorithm to extract a hierarchy of clusters; this hierarchy will represent the structures apparent at different scales in the spatio-temporal events database.

3.1 Recursive Clustering

The use of $BigData$ Clustering is a very powerful approach to extract knowledge about human activities in order to help decision making in big areas [9].

There are several clustering algorithms that produce a hierarchical structure of point samples, Amoeba [8], Chameleon [15], OPTICS [1] and HDBSCAN [22] are examples of such algorithms. In general, hierarchical clustering produces a complete hierarchy: from a single cluster containing all observations to separate clusters for each observation. In practice, most applications using this family of algorithms will cut the hierarchy trough a threshold value (a scale) to obtain a flat representation [22].

In this paper we propose a different approach, instead of using an algorithm that produces a whole hierarchy of clusters, we will use an algorithm that produces a flat representation and then recursively apply it to each cluster found. For this purpose we will use the algorithm DBSCAN [7], which finds density based clusters on databases with noise (that is, observations that do not belong to any cluster).

The decision to use DBSCAN is based on the following remarks:

- The signal coming from the event database will contain samples that represent regular activity and samples that are just the outcome of random processes in the study area (in the case of the geolocated Twitter feed, the latter could be interpreted as the noise commonly reported in Twitter). In this case, the ability of DBSCAN to detect clusters in noisy samples is of importance.
- Regular activity clusters are arbitrarily shaped. There is no reason to assume a particular shape for the regular activity zones, this makes the use of DBSCAN more appropriate to detect the shape of the regular activity zones than K-Means or any other Voronoi based tessellation, since these latter assume that the clusters are convex.
- Cluster algorithms that return a hierarchical structure are generally geared towards finding the most relevant structures in the data, regardless of scale [22]. This means that the flat representation needed to evaluate regular

activity will only contain clusters that are significant along a wide range of scales. In the case of the technique we are proposing here, it is important to have clusters representative of each scale.

The main disadvantage of using DBSCAN is the introduction of the parameters eps_0 and $MinPoints$. For each iteration of the recursive clustering algorithm it is necessary to set appropriate values for these parameters, and this can only be done heuristically and not in a fully tractable way. The heuristic proposed in [7] for determinating suitable eps_0 and $MinPoints$ consists on the examination of the *Sorted K-distance graph* [2]. This limitation will be further discussed in sections 4 and 5.

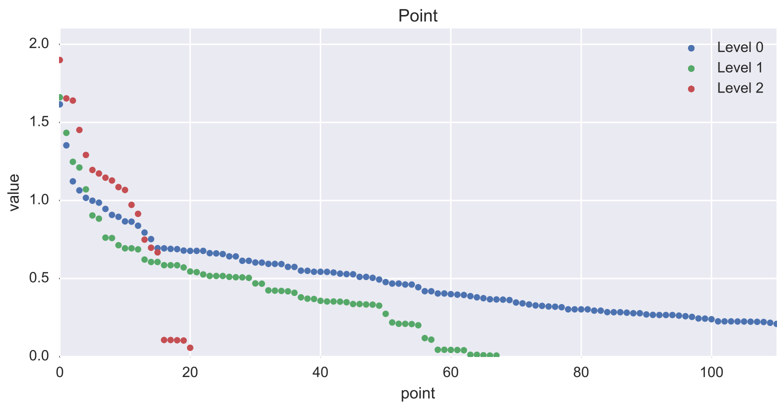


Fig. 1. K-distance graphs for three successive cluster levels. The distance value of the "valley" used to estimate eps_0 drops with each iteration.

On each iteration of the recursive DBSCAN procedure, the eps_0 value passed to the DBSCAN implementation is multiplied by a scaling factor (*decay*), this parameter represents the drop in the relative density of the clusters detected at each scale. Figure 1, shows the different *k-distance graphs* for each scale level for a sample from the experiment that will be discussed in Section 4. As can be seen from the graph, the distance of the "valley" drops as the scale increases (at greater resolution levels), this implies that we must use smaller eps_0 values at each iteration.

The process of regular activity patterns extraction begins, as in [21] and [11], by dividing each day in arbitrary time segments. Ideally, each of these intervals would represent periods where the processes producing the data are stable, for example, the morning commuting peak, working hours, etc. The general workflow

[2] The *Sorted K-distance graph* plots the number of points that exhibit a given distance to their first k-neighbors.

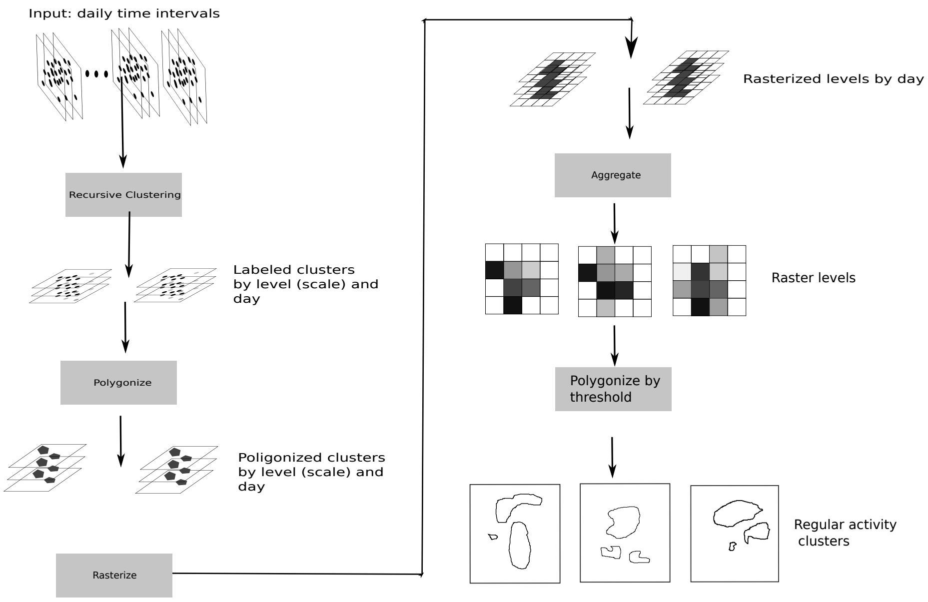


Fig. 2. Diagram showing the workflow for extracting regular activity polygons across multiple scales.

of the process to build the regular activity patterns across the scale levels is shown in Figure 2.

The recursive clustering procedure returns a list of labeled points, the labels represent the clusters to which each point belongs. The next step in the procedure is transforming these sets of labeled points into polygons for each cluster. In order to preserve the property of DBSCAN of producing arbitrarily shaped clusters, we polygonize the points using the Alpha Shape [6] instead of the Convex Hull. It is important to note that, although in general, Alpha Shape extraction involves a parameter determining how closely the polygons follow the underlying points, it is possible to extract an optimal alpha value subject to the following restrictions: 1) the number of connected components is given; 2) all points are either on the boundary or in the interior of the regularized version of the alpha-shape (no singular edges). The first condition implies that each cluster will be represented by a single polygon and the second one that such polygons will be simple.

Once the polygons for each day (and time segment) have been extracted, the next step is to "average" those polygons to find the regular zones of activity. To do this, the polygons are rasterized, i.e. converted to an image whose pixel values are 1 if the pixel lies within the polygon and 0 otherwise. This rasterization introduces another parameter, the *resolution*, that is, the pixel size in the rasterization process.

The rasterized polygons are then aggregated over the whole study period, thus obtaining images whose pixel values represent the number of days a given pixel has been inside a cluster. Figure 6 shows examples of such images for the dataset used in the experiment discussed in section 4.

The images obtained represent the activity patterns of the spatio-temporal events at different resolution levels. Now, in order to characterize this patterns we need a way of assigning the characteristics of the underlying point distribution to the patterns extracted, in the same fashion as Lee, et al. characterize each Voronoi polygon with the count, diversity and movement variables or Frias-Martinez, et al., use the point count aggregated over twenty minute intervals.

In order to achieve this characterization, the raster images are polygonized. This is done by cutting by a threshold value. This threshold represents the number of days a pixel must belong to a cluster in order for it to be considered within zone of regular activity.

After this process, the zones of regular activity are obtained as sets of polygons for each resolution (scale) level. The final step is the characterization of these polygons. This characterization is not considered a part of the regular activity zones extraction since it depends on the objectives of the analysis and the general characteristics of the processes producing the events database.

Before moving on to the experiment that will demonstrate an application of the technique we are proposing, we will briefly discuss all the free parameters involved in this procedure:

eps_0 parameter: Threshold distance above which samples are considered noise in DBSCAN. The heuristic is to find the first "Valley" in the *k-distance graph*.

$MinPoints$ parameter: Number of points at which the eps_0 value crosses the *k-distance graph*. The heuristic is the Same as eps_0.

$decay$ parameter: Drop in eps_0 value as the resolution increases. The heuristic is the observation of the different *k-distance graphs* for successive resolution levels.

$MinSamples$ parameter: Minimum number of points in a cluster to be a candidate for recursive clustering. The heuristic is that below this cluster size, there will be no further resolution levels, so this is should be set according to the minimum scale of detected activity clusters.

$resolution$ parameter: Pixel size for the rasterization of each scale level where the size of the detected regular activity zones will be around 4 times the $resolution$ size.

$threshold$ parameter: Proportion of days a pixel must belong to a cluster to be considered part of the regular activity patterns. The heuristic explanation is that at larger $threshold$ values the resulting polygons will be present in more individual (daily) samples, so this parameter should be set according to a statistical definition of the usual activity.

In the following section we will describe an experiment showing an application of the proposed technique to the extraction of the regular patterns of activity, using the geolocated Twitter feed, for the Central region of Mexico.

4 Experiments

As a first use case of the proposed technique, we will extract the regular patterns of activity in the Central Mexico area. The test database consists of all geolocated tweets from October 10 2014 to April 4 2015, there are 5,415,827 tweets within this period. Prior to the construction of the base scenario, we need to deal with the *pollution* commonly encountered in the tweeter feed [[14], [20]], this means that we must perform some preprocessing to clean up the database. In this case we want to filter out tweets by users that have more than one update within 100 meters of their original location in the same time period. The rationale behind this filtering is that this kind of behavior might be representative of bots or that it might artificially alter the shape of the clusters without representing the regular activity of the population.

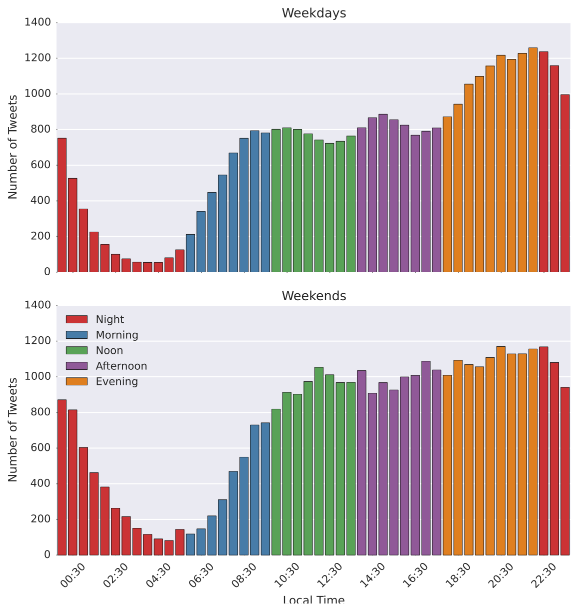


Fig. 3. Twitter activity for each time segment. The bars represent the amount of geolocated tweets aggregated over 30 minutes intervals.

The next step in the construction of the base scenario is the segmentation of time, in this case we will use intervals commonly used in several urban activity studies. Different regular activity patterns will be built for weekdays and weekends, since the patterns of activity are expected to be different. Each day will be segmented as follows: **Morning**: From 06:00 to 10:00, **Noon**: From 10:01 to 14:00, **Afternoon**: From 14:01 to 18:00, **Evening**: From 18:01 to 22:00

and **Night**: From 22:01 to 06:00.The resulting temporal activity patterns can be seen in Fig. 3.

Once we have defined the time segments, it is necessary to set appropriate values to the parameters defined in section 3. At this stage the significant parameters are eps_0, *decay* and $MinPoints$, since these will determine the number of scale levels and the clusters detected at each level. Following [7], it is possible to find suitable eps_0 and $MinPoints$ values by examining the *sorted k-distance* graph for the events population. Fig. 4 shows the *sorted k-distance* graphs for the different time intervals defined above (the Night interval is left out since it has very little activity).

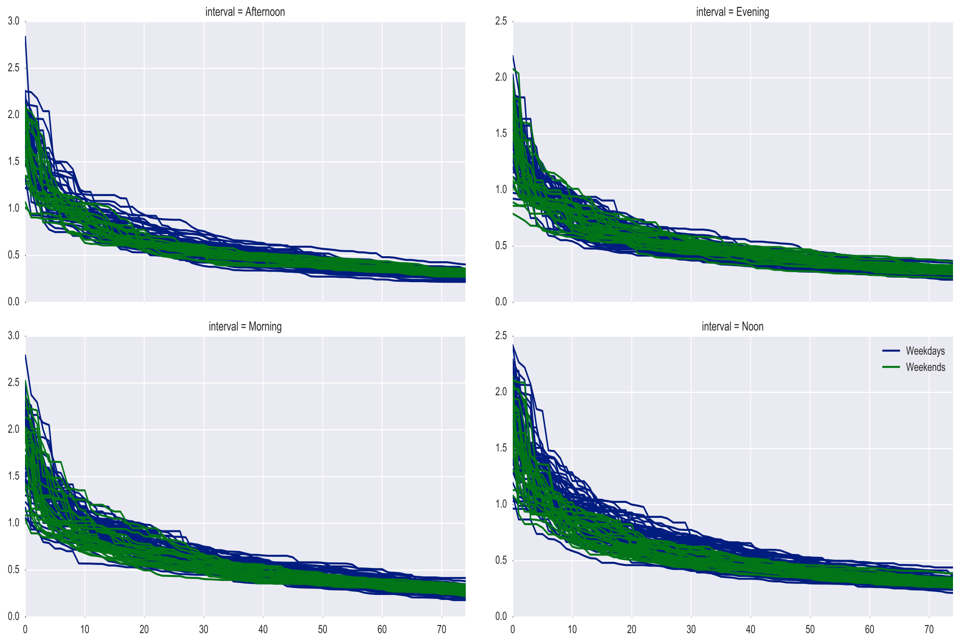


Fig. 4. Sorted k-distance plots for all time segments and every day in the experiment.

According to the heuristic proposed in [7], the value of eps_0 can be determined by finding the first "valley" in the *sorted k-distance* graph, that is, the distance at which we observe a sharp change in the decay rate of the nearest neighbor distance vs. the number of points. In the case of the present technique, we must find a value that is a suitable candidate for clustering every interval of every day within the study period. In Figure 5, we show a single interval *sorted k-distance* graph. As can be seen from the graph, every day shows a different "valley", so we end up having a range of suitable eps_0 values, which makes the selection somewhat arbitrary. Although this might seem a major obstacle, we will show that by choosing a value of eps_0 (and implicitly selecting $MinPoints$) around the middle of the aforementioned range, we can find average (in the sense described in section 3) activity clusters that resemble the known spatio-temporal patterns of activity in the study area. For the rest of the present experiment we will use a value of 0.8 for eps_0 and 18 for $MinPoints$.

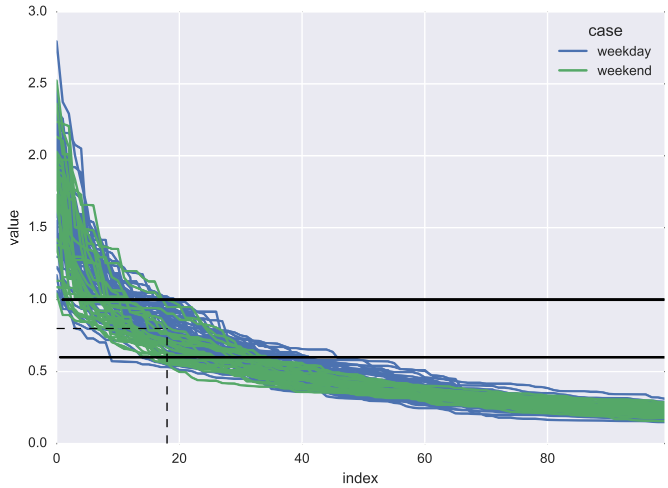
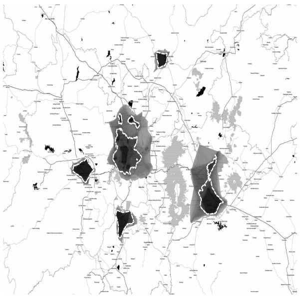


Fig. 5. Sorted k-distance plots for a single time segment. The horizontal lines show the range of "valley" candidates and the dashed lines show the selected values for eps_0 and $MinPoints$.

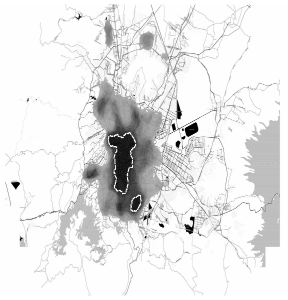
The next parameter we need to choose, as described in section 3, is $MinSamples$, the minimum number of points a cluster must have to be considered a candidate for having clusters of bigger scale. In the case of $MinSamples$, the selection criteria is based on the cluster size, in number of points, at the maximum resolution level, in the case of the current dataset we will set this value at 50.

The final parameter we need in order to perform recursive clustering with DBSCAN, is the *decay* value. This parameter represents the drop in the distance that separates the cluster samples from the noise samples in each DBSCAN iteration. As can be seen from Figure 1, the distance value of the "valley" drops from around 0.6 in the top level to around 0.1 in the bottom level, since the decay rate in the algorithm is constant we will use a *decay* value of 0.4 which represent a 0.8 drop across two levels (this value is set after examination of several point samples).

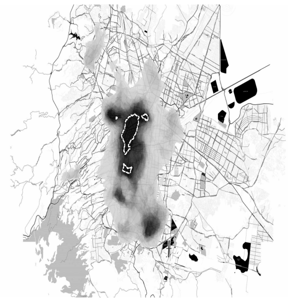
Finally, we need to set values for the *resolution* and *threshold* parameters. For the *resolution* parameter we will use a value o 100 meters, which means the the smaller cluster size we will detect as regular activity will be similar to a 400m by 400m square. The *threshod* value will be set at 0.75 which means that



Evening Level 0.

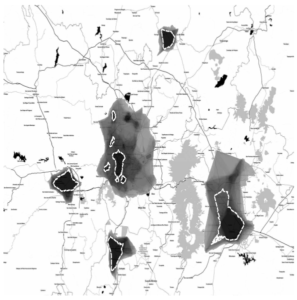


Evening Level 1.

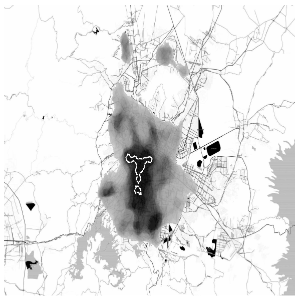


Evening Level 2.

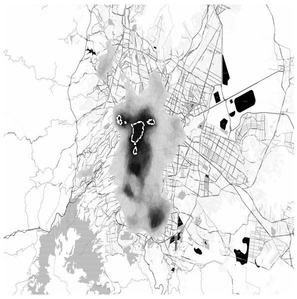
Evening Level 3.



Afternoon Level 0.



Afternoon Level 1.



Afternoon Level 2.

Afternoon Level 3.

Fig. 6. Aggregated activity rasters, with threshold cut polygons in dashed lines, for the Evening and Afternoon periods.

we require a cluster to be present in at least 75% of the sample in order to be considered a regular activity zone.

With all the parameters set, the regular activity zones are calculated, the resulting aggregated rasters are shown in Figure 6, together with the *threshold* cut resulting polygons.

4.1 Discussion of the Results

At the smaller scale (Level 0 in Figure 6), our technique is able to detect the greater metropolitan areas of Mexico City, Puebla, Toluca, Pachuca and Cuernavaca, in the Central Mexico region. As we increase the resolution, the sample density in the smaller cities (Puebla, Toluca, Pachuca and Cuernavaca), does not allow for the recursive algorithm to detect larger scale activity, the opposite is true for Mexico city, where we are able to detect up to three scales within the city.

By comparing the patterns found for the Afternoon and Evening intervals, we see that in the latter, the activity is more dispersed and Level 0 shows activity peaks in the northern low income housing suburbs. On the other hand, the activity for the Afternoon segment is more concentrated around the Central Buissnes District (CBD) of the city. This results are consistent with the known activity patterns for Mexico City. For example, Suarez and Delgado [25] performed a study in the Job-Housing ratio and found the same T-shaped pattern for the CBD. From the same study, we can see that the job to housing ratio of the northern low income suburbs is very low, which means it is mostly a residential area, this is in line with our results that show activity peaks for those areas only at the Evening intervals, that is when people are mostly at home.

5 Conclusions and Further Work

The technique presented in this paper represents an improvement on the available methods for determining the regular activity zones within the geolocated Twitter feed. Its main improvements are: 1) the ability to detect regular patters at different scale or resolution levels. This allows us to detect both major urban areas and activity zones within those areas that have a high enough activity density, 2) Using the Alpha Shapes to polygonize the clusters, allows us to account for the shape of the regular activity zones. This represents an improvement compared to the use of Voronoi tessellations.

The qualitative analysis of the obtained regular activity zones, show great accordance with the known activity patterns for the study area, mainly with the spatial distribution of the Job-Housing ratio. Albeit a formal quantitative validation of the activity patterns is missing, the fact that the technique is able to reproduce qualitatively the spatial distribution of activities within the city is very promising. The next step is the quantitative validation of the results obtained. For this, an updated Job-Housing ratio map must be built and the mobility patterns could be extracted from an Origin-Destination Survey (although the must recent one available for Mexico City is from 2007, the government has announced plans to conduct a new study).

It is also necessary to find more tractable approximations to setting values for the parameters used in regular activity extraction. Ground truthing against measured activity distributions would provide basis for a calibration-validation approach. Finally, the multi-scale regular activity zones could be used to detect unusual crowd activity at various scales. The rationale behind this is that unusual events also exhibit scale differences. For example, it is known that important large scale events, such as the Super Bowl or the Arab Spring, produce a general increase of messages in the social networks, while localized small scale occurrences, such as festivals, demonstrations or accidents, produce small clusters of messages around the locations affected.

References

1. Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander, J.: OPTICS: Ordering Points to Identify the Clustering Structure. In: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data. pp. 49–60. SIGMOD '99, ACM, New York, NY, USA (1999)
2. Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, P., Wang, J., Batty, M.: Hierarchical organisation of Britain through percolation theory. arXiv:1504.08318 [physics] (Apr 2015)
3. Atefeh, F., Khreich, W.: A Survey of Techniques for Event Detection in Twitter. Computational Intelligence 31(1), 132–164 (Feb 2015)
4. Boettcher, A., Lee, D.: EventRadar: A Real-Time Local Event Detection Scheme Using Twitter Stream. pp. 358–367. IEEE (Nov 2012)
5. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: #Earthquake: Twitter as a Distributed Sensor System. Transactions in GIS 17(1), 124–147 (Feb 2013)
6. Edelsbrunner, H.: Smooth surfaces for multi-scale shape representation. In: Thiagarajan, P.S. (ed.) Foundations of Software Technology and Theoretical Computer Science, pp. 391–412. No. 1026 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Dec 1995)
7. Ester, M., Kriegel, H.p., S, J., Xu, X.: 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 (KDD-96). pp. 226–231. AAAI Press (1996)
8. Estivill-Castro, V., Lee, I.: Amoeba: Hierarchical Clustering Based On Spatial Proximity Using Delaunaty Diagram (2000)
9. Estrada, R., Molina Villegas, A., Perez-Espinosa, A., Reyes-C, A., Quiroz, J., BravoG, E.: Zonification of heavy traffic in mexico city. In: Proceedings of the International Conference on Data Mining (DMIN). p. 40. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2016)
10. Frias-Martinez, V., Soto, V., Hohwald, H., Frias-Martinez, E.: Characterizing Urban Landscapes Using Geolocated Tweets. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). pp. 239–248 (Sep 2012)

11. Frias-Martinez, V., Frias-Martinez, E.: Spectral clustering for sensing urban land use using Twitter activity. Engineering Applications of Artificial Intelligence 35, 237–245 (2014)
12. Fujisaka, T., Lee, R., Sumiya, K.: Detection of Unusually Crowded Places through Micro-Blogging Sites. In: 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA). pp. 467–472 (Apr 2010)
13. Gabrielli, L., Rinzivillo, S., Ronzano, F., Villatoro, D.: From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns. In: Nin, J., Villatoro, D. (eds.) Citizen in Sensor Networks, pp. 26–35. Lecture Notes in Computer Science, Springer International Publishing (Jan 2014)
14. Hurlock, J., Wilson, M.L.: Searching Twitter: Separating the Tweet from the Chaff. In: ICWSM. pp. 161–168 (2011)
15. Karypis, G., Han, E.H., Kumar, V.: Chameleon: hierarchical clustering using dynamic modeling. Computer 32(8), 68–75 (Aug 1999)
16. Khan, A.: The role social of media and modern technology in arabs spring. Far East Journal of Psychology and Business 7 No 1 Paper 4 April(4), 56–63 (2012)
17. Kim, T., Huerta-Canepa, G., Park, J., Hyun, S., Lee, D.: What's Happening: Finding Spontaneous User Clusters Nearby Using Twitter. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom) (Oct 2011)
18. Kisilevich, S., Mansmann, F., Nanni, M., Rinzivillo, S.: Spatio-temporal clustering. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 855–874. Springer US (Jan 2010)
19. Kohonen, T.: The self-organizing map. Proceedings of the IEEE 78(9), 1464–1480 (Sep 1990)
20. Lee, R., Sumiya, K.: Measuring Geographical Regularities of Crowd Behaviors for Twitter-based Geo-social Event Detection. In: Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. pp. 1–10. LBSN '10, ACM, New York, NY, USA (2010)
21. Lee, R., Wakamiya, S., Sumiya, K.: Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web 14(4), 321–349 (2011)
22. Li, L., Xi, Y.: Research on Clustering Algorithm and Its Parallelization Strategy. In: 2011 International Conference on Computational and Information Sciences (ICCIS). pp. 325–328 (Oct 2011)
23. Oh, O., Agrawal, M., Rao, H.R.: Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. Information Systems Frontiers 13(1), 33–43 (Sep 2010)
24. Starbird, K., Palen, L., Hughes, A.L., Vieweg, S.: Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. pp. 241–250. CSCW '10, ACM, New York, NY, USA (2010)
25. Suarez, M., Delgado, J.: Is Mexico City Polycentric? A Trip Attraction Capacity Approach. Urban Studies 46(10), 2187–2211 (Sep 2009)