Chapter 1

Geomatics Applications to Contemporary Social and Environmental Problems in Mexico

Jose Luis Silván-Cárdenas, Rodrigo Tapia-McClung,
Camilo Caudillo-Cos, Pablo López-Ramírez,
Oscar Sanchez-Sórdia and
Daniela Moctezuma-Ochoa

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/64355

Abstract

Trends in geospatial technologies have led to the development of new powerful analysis and representation techniques that involve processing of massive datasets, some unstructured, some acquired from ubiquitous sources, and some others from remotely located sensors of different kinds, all of which complement the structured information produced on a regular basis by governmental and international agencies. In this chapter, we provide both an extensive revision of such techniques and an insight of the applications of some of these techniques in various study cases in Mexico for various scales of analysis: from regional migration flows of highly qualified people at the country level and the spatio-temporal analysis of unstructured information in geotagged tweets for sentiment assessment, to more local applications of participatory cartography for policy definitions jointly between local authorities and citizens, and an automated method for three dimensional (3D) modelling and visualisation of forest inventorying with laser scanner technology.

Keywords: crowdsourcing, airborne laser scanner, crime analysis, migration, volunteered geographic information

1. Introduction

The term geomatics was originally conceived by Michel Paradis, a French-Canadian surveyor, as the discipline of gathering, storing, processing and delivering spatially referenced

information [1]; as such, geomatics has been tied to the development of geospatial technology since its birth. The Encyclopaedia of Geographic Information Science by Karen Kemp defines geomatics as the 'science of building efficient Earth related data production workflow' [2]. According to this definition, the discipline of geomatics 'truly highlights the necessary shift from a technology-oriented silo approach to a data-flow-oriented system approach geared toward a result in a given context' [2]. It is the result-oriented mode that stresses the need for a transdisciplinary approach, which has been adopted by researchers at the Geography and Geomatics Research Centre in Mexico (CentroGeo).

As the technology evolves, the research field of geomatics has to necessarily expand along its entire workflow, from data acquisition to geospatial information dissemination. For instance, the georeferencing capability of mobile devices and their extensive use in social networking are producing unprecedented amounts of information that can be of high relevance for many important topics such as security, marketing, mental health, disaster management, etc. Consequently, social media analysis is becoming a very important research topic within geomatics and its related fields.

In Section 2, we provide a brief review of major steps within the geomatics approach, from data acquisition processes and processing techniques to the analysis and visualisation methods used for information extraction and representation. Then, in Section 3 we provide illustrative examples of applied geomatics research to contemporary social and environmental problems in Mexico. Section 4 ends this chapter by providing some concluding remarks.

2. The geomatics approach

In this section, we discuss the general steps involved in addressing social or environmental issues from geomatics. The goal is to make a general review of data acquisition, processing, analysis, visualisation and interpretation, providing examples from different fields such as remote sensing, crime analysis or social media.

2.1. Data acquisition processes

2.1.1. Remote sensing

Since Gaspard-Félix Tournachon took the first aerial photograph in 1858 from a tethered balloon over Paris, the interest for observing the Earth from afar has grown to the point that cameras are put on board of any sort of flying devices including kites, balloons, airplanes, rockets, satellites, spatial stations and unmanned aerial vehicles (drones). Indeed, aerial photography has been the most common, versatile and economical form of remote sensing, but other types of sensors besides cameras have also been developed [3].

In this sense, remote sensing is a continuously evolving field that is devoted to the design and development of new and effective techniques for data acquisition of the Earth's surface from remote locations, typically from space and aircrafts. All these techniques share a common

http://dx.doi.org/10.5772/64355

principle: to record the energy, typically the electromagnetic radiation, that has interacted with the Earth's surface in order to retrieve some information about it.

The range of frequencies (or wavelengths) of the electromagnetic radiation that the sensor is sensitive to is of prime importance because it determines which materials can be detected. It also influences whether to use the natural illumination of the sun or to use an artificial energy source. Sensors are active or passive depending on whether they include an artificial source of energy or not. Thus, for instance, infrared and thermal cameras are considered passive sensors because they sense the reflected near-infrared light and the emitted thermal infrared from hot bodies, respectively, whereas radar and lidar systems are considered active sensors because they send microwave and laser beams, respectively, and detect the backscattered energy.

The ability to measure quantities of radiant energy (radiance/reflectance, emittance, backscattering, etc.) would have not been as useful as it is, except because the sensors are coupled with global positioning systems (GPSs) and inertial measuring units (IMUs) for measuring location and orientation, thus enabling the production of digital representations of surface features that can be integrated into geodatabases.

Furthermore, a substantial body of knowledge from related fields, such as radiative transfer theory, imaging spectroscopy, image/signal processing and computer vision, has been advanced that allows deriving ready-to-use information in the form of data layers that can be overlaid within a geographic information system (GIS). These layers include vegetation indices, digital elevation models, surface temperature, soil moisture, rainfall, snow cover, night light, impervious surface, mineral abundance and land-cover types, to name just a few. These surface features are specified by the various resolutions and dynamic ranges of the sensor (spatial, temporal, spectral and radiometric). The former refers to the smallest spatial, temporal, spectral and radiometric difference, which the sensor can resolve, whereas the latter refers to the largest differences that can be resolved. Hence, depending on the resolution/dynamic-range characteristics of sensors, they have distinct uses.

2.1.2. In situ data collection

In situ data collection refers to the collection of georeferenced data (mainly points and areas) measured on the ground for a number of reasons, such as validating cartographic or remotely-sensed products, producing data layers, model calibration and/or validation, or simply gaining some understanding of the study area, amongst other reasons.

Regarding the methods for in situ data collection, one can guess that there are as many as the fields involved. One fundamental question to answer before anything is done is: What do we need to know from the ground? Then, we can decide the variables to be measured, the sampling scheme and personnel and instrumentation needed. Among the many decisions to make is whether to perform a random or systematic sampling; whilst the former is preferred for accuracy assessment purposes, the latter is desirable for spatial analysis, for example, spatial interpolation.

Today, there is a growing number of affordable digital technologies that enable the collection and real-time analysis of georeferenced field data. Not only is the increase in performance, resolution and portability of measuring devices but also the functionality that enables on-site analysis and visualisation that is making the in situ data collection more efficient with reduced uncertainty [4]. Laser-based technology (e.g. range finders, dendrometers, terrain profilers, terrestrial laser scanners, etc.) has enabled the measurement of inaccessible locations and generation of coloured point clouds that capture the three dimensional (3D) structure of the sampled site. On the other hand, modern communication protocols, mobile device network coverage and cloud storage capabilities are also facilitating field data management and sharing in unprecedented ways.

2.1.3. Crowdsourcing

The ubiquitous use of mobile devices and Internet access has fostered the ability of citizens to collect their own data for varied purposes. Many apps and platforms have been developed that allow citizens to collect data. GeoKey is a backend platform that allows the creation of customised projects [5]. One still needs to programme a frontend, but it is quite versatile in the types of data it can handle. GeoCitizen is a platform developed for community-based spatial planning. Its goal is to provide means and information for citizens to access data and get involved in every step of the planning process [6]. Ushahidi is a well-known platform used for crisis mapping [7]. It gained momentum during and after the massive earthquake that hit Haiti in 2010 [8]. OpenTreeMap allows users to collaborate in creating a massive inventory of trees that are useful for ecosystem management and urban forestry [9]. iNaturalist focuses on users collecting data about observations of the natural world [10]. Waze has also become a very common platform that allows real-time communication with other users reporting traffic conditions whilst driving [11]. NoiseTube has also been used for participatory noise pollution mapping and monitoring [12].

Without necessarily challenging the existence of official records, it is increasingly common to compare what the official figures tell with what the citizenry observes and experiences on its everyday life.

Crowdsourcing and volunteered geographic information (VGI) are two terms that have been more pervasive in the academic literature. But what, if any, is the difference between them? Crowdsourcing can be found in many different topics, not just geographical information and 'implies a coordinated bottom-up grassroots effort to contribute information' [13]. For some, VGI represents an 'unprecedented shift in the content, characteristics, and modes of geographic information, creation, sharing, dissemination and use' [14]. Others, such as Harvey, propose that not all crowdsourced data are volunteer data. He suggests making a distinction when data are collected with an 'opt-in' or an 'opt-out' agreement [15].

Nonetheless, both ideas—crowdsourcing and VGI—rely on data being contributed by many users. In a sense, they are strong advocates of the 'wisdom of the crowds' and collective intelligence: the idea of whether a product created collectively is better than the best individual product [16, 17].

http://dx.doi.org/10.5772/64355

The deluge of mobile apps makes it possible to crowdsource data practically anywhere. In Mexico, however, strong biases can be introduced with this form of data collection, as it may be far more popular in urban settings with the added issue that not all regions in the country have the same mobile network coverage [18].

2.2. Processing techniques

Data-processing techniques refer to techniques for data preparation prior to any information extraction. These techniques include data reformatting, cleaning, rectification, denoising, enhancement, etc. Although a thorough review of such techniques is beyond the scope of this chapter, it is worth noting that most techniques that operate in raster formats come from the digital image-processing field, where theoretical developments have been around filtering techniques in both the space and frequency domains. Additionally, techniques such as principal components analysis (PCA) and minimum noise fraction (MNF) are applied as spectral transformations of multispectral and hyperspectral images, whilst some spatial, multiscale representations, for example, wavelets, are used for image denoising or spatial enhancement (pansharpening).

In fields such as crowdsourcing or social media analysis, the preprocessing can be even more important (since there is no adequate way to calibrate the 'instruments' used to acquire data), but, opposed to remote sensing, there is no sound theoretical framework from where to draw techniques. This situation requires, in the best case, the use of some form of ground truthing to discard spurious data. Wherever reliable data are not available, the researcher must resort to his/her domain knowledge or heuristic algorithms to preprocess the data.

2.3. Analysis and interpretation

The increasing production of spatial data from both official and non-official sources and with unstructured formats has placed a larger complexity in its management and analysis. On the one side, information granularity has incremented both spatially and temporally, thus making it necessary to develop analytical tools that simultaneously take into account space and time for decision-making. On the other side, the great diversity of sources of information that share the spatial component has triggered the efforts for interoperability, which implies the possibility of combining multidimensional information that can provide potential knowledge. In this section, we describe some of the most pervasive methods of analysis used by geospatial technologies.

2.3.1. Cluster analysis

Generally speaking, cluster analysis refers to the process of grouping objects into classes by some measure of similarity. These objects can be either abstract, as the companies in the stock market, or physical as the states within a country. The similarity measures used on cluster analysis depend on the kind of objects and the characteristics being analysed. If the interest is in grouping earthquake occurrences, then the Euclidian distance is a reasonable similarity

measure, but if we are grouping counties around some measurement of its economic performance, the Mahalanobis metric could be a reasonable choice.

Cluster analysis has been successfully used in many applications: market research uses segmentation to target products; in biology, it is used for taxonomy and DNA sequencing; in image recognition, it is used in image segmentation.

Certainly, cluster analysis is not new within the field of geographic data analysis; ISODATA has been in use for over 40 years in multispectral image classification [19]; the famous John Snow map of the cholera outbreak in London is also a case of cluster analysis, and the concept of regionalisation, when approached from a spatial analysis perspective, can be interpreted as a case of geographically constrained clustering, that is, clusters in which observations are grouped together by their similarity in the feature space but restricted to their neighbourhood relations in the geographical space [20].

Recently, the increase in the quantity of data collected every day from a great number of disparate sources has stemmed a new interest in the techniques derived from cluster analysis. One of the reasons of this recent interest lies in the flexibility of the similarity measures that can be used. This is especially important when working with what has been labelled as unmodelled data, that is, data that are not structured for analysis, such as natural language. This kind of information has become more frequent as technologies such as social media and the pervasiveness of sensors are becoming commonplace.

Although there are cluster analysis techniques that clearly come from the statistical modelling tradition, such as the work of Kulldorf on epidemics or ISODATA [21], the recent increase in clustering methods comes from the algorithmic culture. Applications such as handwritten recognition or image segmenting make extensive use of clustering methods from the algorithmic culture [22–25].

In the field of geographic data analysis, there are also some important developments. In particular, the field of geographic knowledge discovery (GKD) is gaining recognition as is evident from the amount of conferences and special issues devoted to the topic ([26, 27], amongst others).

On the subject of cluster analysis as a mean for extracting geographic knowledge from unmodelled data, there have been some interesting recent developments. Frias-Martinez et al. proposed a technique for extracting land-use information from geolocated Twitter feed and used spectral clustering for the extraction of regular activity zones [28]. Lee et al. used *k*-means clustering to detect unusual crowds also using geolocated tweets. These works rely solely on the spatio-temporal properties of the data, which is interesting because the techniques developed could be easily translated to work with different datasets, such as mobile telephone records [29].

There are also some interesting examples that combine the spatio-temporal properties with the semantic content of the messages. Amongst these, we find the work of Gabrielli et al. who deduced trajectories from the geolocated Twitter feed and enriched these trajectories with semantic information from the users (e.g. whether the user is a tourist) and the surroundings

(the types of venues located around the user at a given moment) [30]. Also, the works of Boetcher and Lee or Kim et al. present techniques for the detection of local clusters of activity around specific topics of interest [31, 32].

This development in the GKD field, from an academic perspective, has happened in parallel with the development of the data-mining field in the application-driven environment of start-ups and technology corporations. Currently, as the academic field matures, it is beginning to catch up with the technology side developed in the commercial world. The shift of focus towards real-time analysis [33] stresses the need to not only develop better algorithms but also develop them on top of a technological stack that allows the scaling up needed to solve the problems associated with real- or near-real-time analysis.

In the GIS field, the recent development of the CyberGIS paradigm attempts to build a bridge between traditional GIS and new advances on distributed data stores, parallel computing and collaborative workflows [34–36]. Research on the parallelisation of *k*-means and the application of the map-reduce programming paradigm to cluster analysis in general are examples of the direction of technology research within the field of cluster analysis in a GKD framework [37, 38].

2.3.2. Network analysis

Network analysis in the geospatial community generally refers to analysis techniques associated with the optimisation of transportation routes. In this section, we investigate techniques that originated in the field of graph theory to analyse social networks, applied to geographical phenomena—particularly, migration flows.

Migration between metropolitan areas can be conceived as a weighted graph in which nodes (n) are the cities and the edges (m) are the flows between them. In transport analysis literature, there are several techniques to deal with networks; one of the most frequently cited is the nodal region approach [39]. This method is used for quantifying the degree of association between pairs of cities in a way that allows the identification of the strongest association of the network. The result is a graph with a maximum of $(n - 1)$ edges. Further modifications were introduced by Graizbord [40] and Suárez and Delgado [41] in order to provide more flexibility in the hierarchy of the nodes and the size of the filtered graph, as well as some restrictions in the definition of salient flows, such as the comparison from a gravitational model or previous data on migration flows.

Bender-deMoll mentions in his network analysis and mapping report that characterisation of flows of goods and people is a classic field of application of social network [42]. Networks are used to represent flow patterns between sets of entities and constitute a useful analysis of movement structures. Results of some studies on trade flows have shown to provide more knowledge and have helped predict global resources flow between countries. By analysing data on both forced and voluntary migrations, a strong correlation has been found between the geography and the relationships shown by aggregate flows. In the same way, these flows reflect the social links of migrants, that is, they usually move to places where relatives and/or

friends are located, or to places that information networks have detected to be viable for development.

One way to characterise flows is to detect communities, an exercise similar to cluster analysis. With a binary network, this type of analysis can only be performed if the difference between the number of edges (m) and nodes (n) is not too large. If $m >> n$, edge distribution is so homogeneous that communities do not make any sense. However, community detection is possible if the network is weighted and weights have a heterogeneous distribution [43].

The community detection problem requires partitioning a network in groups of densely connected nodes, where nodes belonging to different communities have disperse links. The quality of resulting partitions is usually measured with the so-called modularity of the partition. The modularity of a partition (Q) is a scalar value between −1 and 1 that measures the density of links inside communities as compared to links between communities. In the particular case of weighted networks,

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta\left(c_i, c_j\right) \tag{1}$$

where A_{ij} represents the weight of the edge between i and j, $k_i = \Sigma_j A_{ij}$ is the sum of the weights of the edges attached to node i, c_i is the community to which node i is assigned, $\delta(c_i c_j)$ equals 1 if c_i and c_j are in the same community and 0 otherwise, and $m = \frac{1}{2}\sum_{ij} A_{ij}$.

The Louvain method to optimise the modularity function finds high modularity partitions on large networks in short time and unfolds complete hierarchical community structures for the network. In the final solution, the output partition contains communities of the most densely linked nodes [44].

2.4. Visualisation and interpretation

Starting around the mid-1990s, geovisualisation—the use of visual representations in order to employ vision to solve spatial problems—entered the GIScience arena. MacEachren et al. provided tools for dynamic exploration of data to help discover relationships and patterns by means of exploratory spatial data analysis (ESDA) [45]. At the turn of the century, the term geovisual analytics started to be heard. It deals with analytical reasoning and decision-making whilst using interactive visual interfaces (e.g. maps and other graphic representations) linked to computational methods and the human capacity of knowledge construction and representation [46]. This section presents some of the most popular visual analytics techniques.

2.4.1. Kernel density

One of the most commonly used hotspot detection methods is kernel density estimation. Its advantages reside in the simple interpretation and its availability in almost any geographical information system [47]. One of this method's weaknesses is the need to accumulate observations in a wide temporal window and unfortunately, as many other hotspot detection methods,

http://dx.doi.org/10.5772/64355

it treats spatial and temporal aspects as separate entities, thus ignoring the spatio-temporal interactions.

2.4.2. Knox's index

Halfway through the twentieth century, Knox proposed a statistical test to detect epidemics [48]. Essentially, it was a statistical independence test for contingency tables classifying individual events that were registered by their location close in time and space. A more robust implementation goes beyond the simple independence test, testing for randomness of the spatial pattern [49]. The null hypothesis is as follows: the occurrence of an event is randomly distributed between the locations. That is, distances in time between pairs of observations are independent to the distances in space. The statistics is as follows:

$$x = \sum_{i=1}^{N} \sum_{j=1}^{i-1} a_{ij}^{\delta} a_{ij}^{\tau} \tag{2}$$

With the following restrictions:

$$a_{ij}^{\delta} = \begin{cases} 1, \text{if the distance between cases } i \text{ and } j < \delta \\ 0, \text{if the distance between cases } i \text{ and } j > \delta \end{cases}$$

$$a_{ij}^{\tau} = \begin{cases} 1, \text{if the distance between cases } i \text{ and } j < \tau \\ 0, \text{if the distance between cases } i \text{ and } j > \tau \end{cases}$$

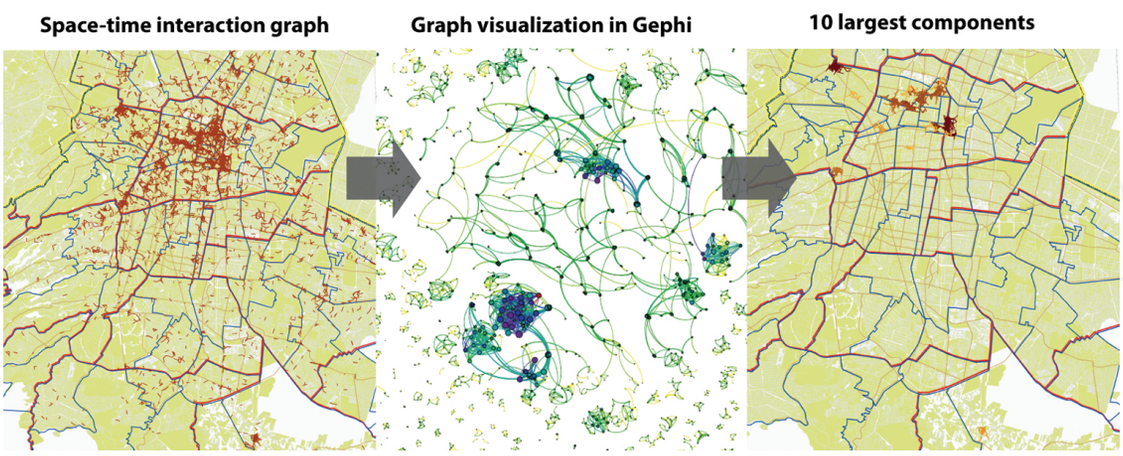

Figure 1. Space-time interaction graph representation and simplification of larceny theft cases in 2009 in Mexico City.

The randomisation technique for the assessment of space-time significance consists on shuffling the temporal distances between cases or events whilst holding the spatial distances constant, and compare the observed and the expected values from Monte Carlo simulations.

The Knox test was originally designed to account for latency periods: time between exposure and the manifestation of symptoms [49].

The added value given to Knox's index by means of a graphic output was to characterise the graph with some simple metrics from network analysis. The only transformation performed on the graph was to invert the role of nodes and edges. The degree of each node and the size of each connected component are useful for detecting significant spatio-temporal events through graph pruning. **Figure 1** illustrates how the application of this index metrics is useful for detecting critical areas in order to design police operations that would align different material and human resources (surveillance cameras, street policemen, police cars, etc.).

2.4.3. Heat maps

Originally designed for displaying financial information that would allow stockbrokers to detect anomalistic behaviours, heat maps were patented, trademarked and made their way into geographical data. Heat maps have been associated to choropleth maps and have become very useful to represent point, line or area density data. Heat maps are also known as density surfaces. They are useful for identifying those areas of a map that have high-density counts within a spatial context [50].

It is probable that after Google released the ability to include heat maps as separate layers using the Maps Javascript API in 2012, the use of heat maps for geospatial data experienced a boom [51]. Since then, many more options have become available.

2.4.4. Flows representation

One of the most often used representations of entities moving between geographical locations is a flow map, in which locations are represented as lines or arrows with their width proportional to the flow magnitudes.

The origin-destination (OD) matrix is an alternative non-geographic visualisation of this kind of data; the magnitudes are represented by the cell colours in a heat map with the rows corresponding to the origins and the columns with the destinations.

A kriskogram is created using a two-step procedure. Firstly, all related geographical units are projected as a set of evenly spaced dots on a straight line called the location line. The order of locations can be arranged using geographical criteria such as the overall orientation of the spatial units, or demographic criteria, such as gross migration or population. In the second step, the migration flow between two places is represented as a half-circle drawn from the origin to the destination in a clockwise direction with the circle's centre located on the middle point between the two corresponding dots on the location line [52].

Flowstrates is an interactive visualisation approach in which the origins and destinations are displayed in two separate maps, and the changes over time of the flow magnitudes are represented in a separate heat map view in the middle [53].

Figure 2 shows examples of the three types of visualisations mentioned in the text. It is evident the kriskogram has two disadvantages: firstly, it loses all spatial reference and secondly it is

impossible to identify the direction of the flow. It facilitates, however, the identification of magnitudes. Heat maps have certain strengths when the network disperses, with few flows. As the network becomes denser, reading it becomes more complex. The method by Boyandin et al. is very interesting since it proposes an interactive exploration tool [53]. Incorporating the heat map allows the identification of trends in migratory flows between pairs of places and avoids information redundancy present in matrix representations by transforming an array of data into one of minimal information in which each flow occupies one row in the heat map. One inconvenience is that as more regions are selected as origins or destinations, the length of the array can grow substantially.

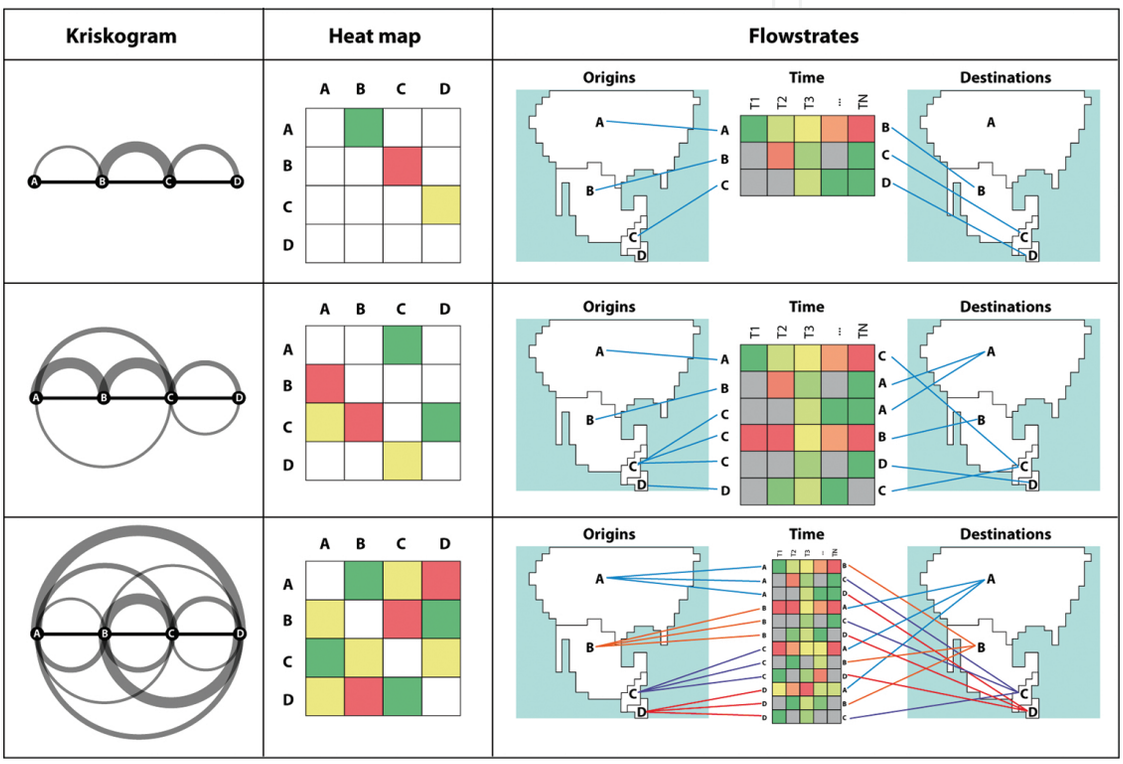

Figure 2. Flow visualisations comparison. Adapted from [52–54].

For our case studies, kriskograms were ruled out because they lose all spatial references. However, we use arcs that avoid overlapping flows. We move away from heat maps in their traditional matrix form and instead use a heat map layer on top of a geographical base. Flowstrates' potential lies in the explicit incorporation of temporal trends. Unfortunately in our case, we lack time series to profit from this representation.

2.4.5. 3D modelling

The development of 3D modelling can be traced back to the 1970s, when efforts of several industries in developing computer-aided design software started. Today, 3D modelling techniques have become an indispensable tool for inventorying and visualisation of objects through digital platforms, but also for producing models with 3D printing devices.

There are several ways for producing 3D scenes. Traditionally, 3D models have been generated manually and algorithmically, especially in the realm of industrial and architectural design. Commercial 3D GIS software, such as ESRI's ArcScene and City Engine, can convert two-dimensional (2D) features into 3D features by applying an extrusion operation (**Figure 3**) and provide extensive libraries of 3D models of vegetation and urban infrastructure [55, 56]. Alternatively, models of actual vegetation and buildings can be generated through remote sensing and computer vision techniques.

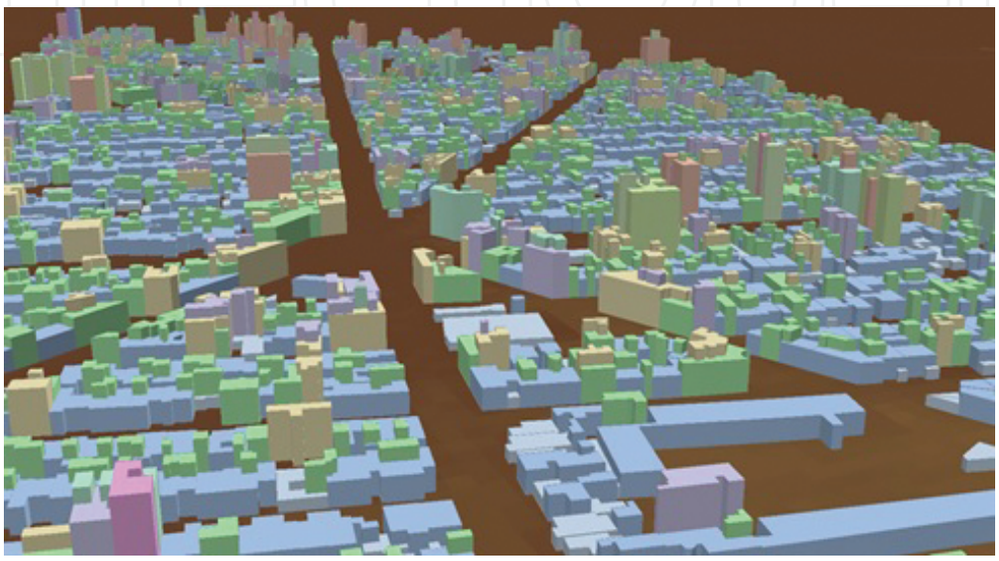

Figure 3. Extruded building footprint from a 2D database.

With the development of laser scanners and advances in photogrammetric techniques, the interest of 3D modelling in the geospatial industry and science has shifted towards the development of new automated or semiautomatic methods for generating photorealistic scenes of the landscape. Close-range data acquisition, such as terrestrial laser scanners (TLSs) and multiple oblique photographs taken with drones, allows the detailed reconstruction of buildings and trees, whereas large-scale projects require the integration of airborne laser scanners (ALSs), aerial photography and satellite-based data acquisitions.

Tree reconstruction and modelling from ALS data have been developed using the voxel approach [36], simple geometrical models such as paraboloids and ellipsoids [57], wrapped surfaces derived by radial basis functions and isosurfaces [58], whereas detailed modelling of trees has been carried out using mobile laser scanners (MLSs), where tree trunk and branches are detected and reconstructed [59]. Buildings are also reconstructed from both laser scanner data [60] and photogrammetric techniques using multiple oblique photographs [61]. These methods are, however, not fully integrated within the 3D GIS platforms but rather are components of remote sensing and photogrammetric-processing systems.

There has also been an increasing demand to use 3D models in virtual reality (VR) and augmented reality (AR) environments, in which virtual and immersive scenes are generated in real time for several applications such as education, training, manufacturing, remote

operations, entertainment, collaborative work, and so on. The key idea is the interaction of humans with 3D models (in place of real objects) that are immersed in a background scene and may include ambient stimuli. Although VR and AR have evolved separately, efforts have been made to integrate these techniques with 3D GIS [62].

The adoption of these technologies has been proved successfully for urban planning, cadastral information updating and for archaeological cultural heritage documentation and visualisation.

2.4.6. Space-time data representations

In the early stages of geographic information sciences, most analyses and representations were focused on static data and models. This is, as Goodchild argues, a consequence of the close relationship that existed within digital data and hard-copy maps [63]. The former was produced by a digitisation of the latter, which implies that digital data had to accommodate to the lengthy and costly procedure of updating, for example, the general topographic maps.

As the field and its associated technology evolved, we have seen an ever-increasing amount of spatio-temporal information gathered: satellite images, GPS traces, climate data, etc. In order to make sense of these data and to fully realise its potential in helping unveil the dynamics of the processes that produce the 'static' patterns observed, we need better tools to digitally represent and analyse spatio-temporal data.

In terms of the digital representation of spatio-temporal data, the early work of Langram and Chrisman on spatio-temporal topology clearly represents a departing point for the evolution of the field [64]. From a theoretical perspective, the work of Hagerstrand on spatial diffusion and space-time geography represents an equally important starting point for space-time modelling from a spatial analysis perspective [65, 66].

Although the field has seen great advances from these early examples, the main issues involving the establishment of the temporal dimension in the GIS field were already present: geographical models need to be explicitly temporal (as Hagerstrand's innovation diffusion [65]), the need of theoretical foundations that explain the way in which the modelled subjects interact in space and time (when studying human populations, this lies within space-time geography, but when we deal with different problems, e.g. ecology, the theories will certainly arise from different fields), and, finally, the need for data structures that allow storing and processing spatio-temporal data in ways that are meaningful to the problems at hand.

3. Case studies

This section presents examples drawn from the experience of the authors working in social and environmental issues, which will help clarify the concepts exposed in the previous sections. Although not always explicit, all of the examples presented here include the steps of data processing, analysis and visualisation as well as results interpretation. The intent is not

to provide a complete explanation of each example but to provide a general application context to complement the general approach presented in the previous sections.

3.1. Social media analysis of subjective well-being

A proposed technique for global polarity classification in short texts, specifically Twitter, is described. The main objective was to obtain a map of subjective well-being for conterminous Mexico; this map will allow us to see the differences in regional perceptions about general well-being. Although this kind of maps can be obtained by traditional methods, such as polls, it is important to note that the amount of resources, human and economic, involved in such exercises, makes it impossible to measure well-being on finer spatio-temporal resolutions. On the other hand, validating a methodology based on social media analysis allows us a very fine-grain analysis, certainly, losing some of the robustness obtained with traditional polling.

For this, we classified the polarity (or sentiment) for each short text (in this case, a tweet). Sentiment analysis is one of the most important tasks in text mining. Nevertheless, this kind of analysis has several challenges related to the complexity of human language, that is, multitude of styles, informal writing, language mixing, short contexts, orthographic and grammatical errors, an always-growing vocabulary, etc. The sentiment classification attempts to determine if one document has a positive, negative or neutral opinion or any level of them (e.g. positive+, negative+, etc.). Determining whether a text document has a positive or a negative opinion is becoming an essential tool for both public and private companies [67]. This tool is useful in knowing 'what people think', which can be important information to help in any decision-making process (for governments, marketing companies, etc.) [68].

3.1.1. Related work

Nowadays, several methods have been proposed in the community of opinion mining and sentiment analysis. Most of these works employ Twitter as a principal input of data and they aim at classifying entire documents as overall positive- or negative-polarity levels (sentiment). Such is the work presented by da Silva et al., which proposes an approach to classify sentiment of tweets by using classifier ensembles and lexicons; tweets are classified as positive or negative. As a result, it is concluded that classifier ensembles formed by several and diverse components are promising for tweet sentiment classification [69]. Moreover, several state-of-the-art techniques were compared in four databases. The best accuracy result reported was around 75%.

Another method for sentiment extraction and classification of unstructured text is proposed by Shahbaz et al. who used five classes: strongly positive, positive, neutral, negative and strongly negative [70]. The proposed solution combines techniques of natural language processing (NLP) at sentence level and algorithms of opinion mining. The accuracy result was 61% for five levels and 75% by reducing to three levels (positive, negative and neutral).

An approach of multi-label sentiment classification was proposed by Liu et al., which has three main components: text segmentation, feature extraction and multi-label classification [71]. The features used included raw segmented words and sentiment features based on three sentiment

dictionaries: DUTSD, NTUSD and HD. Moreover, here, a detailed study of several multi-label classification methods is conducted, in total, 11 state-of-the-art methods have been considered: BR, CC, CLR, HOMER, RAkEL, ECC, MLkNN, and RF-PCT, BRkNN, BRkNN-a and BRkNN-b. These methods were compared in two microblog datasets, and the reported results of all methods are around 0.50 of *F*-measure.

In general, most of the analysed works classify the documents mainly in three polarities: positive, neutral and negative. Moreover, most works analyse social media (mainly Twitter) documents. In this section, we describe a method to classify sentiment in tweets. The sentiment of the messages will be classified into three polarity levels: P (positive), neutral and N (negative). The proposed method is based on several standard techniques such as LDA (Latent Dirichlet Allocation), LSI (Latent Semantic Indexing), term frequency-inverse document frequency (TF-IDF) matrix in combination with the well-known SVM (Support Vector Machine) classifier.

3.1.2. Proposed solution

The overall workflow can be summarised as follows. A preprocessing step is first carried out, then a pseudo-phonetic transformation is applied and, finally, the *q*-gram expansion is generated.

The preprocessing focused on the task of finding a good representation for tweets. Since tweets are full of slang and misspellings, the tweet text is normalised using procedures such as error correction, usage of special tags, part of speech (POS) tagging and negation processing. Error correction consists on reducing words-tokens with invalid duplicate vowels and consonants to valid-standard Spanish words (ruidoooo → ruido; jajajaaa → ja; jijijji → ja). Error correction uses an approach based on a Spanish dictionary, a statistical model for common double letters and heuristic rules for common interjections. In the case of the usage of special tags, twitter's users (i.e. @user) and URLs, they are removed using regular expressions; in addition, 512 popular emoticons were classified into four classes (P, N, NEU, NONE), which are replaced by a polarity tag in the text, for example, positive emoticons such as :), :D are replaced by _POS, and negative emoticons such as :(, :S are replaced by _NEG. Emoticons without any polarity charge are discarded.

In the POS-tagging step, all words are tagged and lemmatised using the Freeling tool for the Spanish language stop words are removed, and only content words (nouns, verbs, adjectives and adverbs), interjections, hashtags and polarity tags are used for data representation [72]. In the negation step, Spanish negation markers are attached to the nearest content word, for example, 'no seguir' is replaced by 'no_seguir', 'no es bueno' is replaced by 'no_bueno', 'sin comida' is replaced by 'no_comida'; a set of heuristic rules for negations are used in this case. Finally, all diacritic and punctuation symbols are also removed.

In a second step, and with the purpose of reducing typos and slangs, a semi-phonetic transformation was applied. Firstly, the following transformations (with precedence from top to bottom) as shown in **Table 1** were carried out.

In this transformation notation, square brackets do not consume symbols and means for any valid symbols. The idea is not to produce a pure phonetic transformation as in Soundex-like algorithms, but try to reduce the number of possible errors in the text. Notice that the last two transformation rules are partially covered by the statistical modelling used for correcting words (explained in the preprocessing step). Nonetheless, this pseudo-phonetic transformation does not follow the statistical rules of the previous preprocessing step.

$\boldsymbol{cx \mid xc \rightarrow x}$	$\boldsymbol{ll \rightarrow y}$	$\boldsymbol{w \rightarrow u}$
$qu \rightarrow k$	$z \rightarrow s$	$v \rightarrow b$
$gue \mid ge \rightarrow je$	$h \rightarrow \in$	$\Psi\Psi \rightarrow \Psi$
$gui \mid gi \rightarrow ji$	$c[a \mid o \mid u] \rightarrow k$	$\Psi\Delta\Psi\Delta \rightarrow \Psi\Delta$
$sh \mid ch \rightarrow x$	$c[e \mid i] \rightarrow s$	

* i denotes the imaginary unit number.

Table 1. List of transformations applied to geotagged tweets.

Finally, along with the bag of words representation (of the normalised text), the four- and five-gram characters of the normalised text were added. Blank spaces were normalised and taken into account to the q-gram expansion; so, some q-grams will be over one word. In addition to these previous steps, several transformations (LSI, LDA and TF-IDF matrix) were conducted to generate several data models for the testing phase.

3.1.3. Results and analysis

For the experiments, a total of 7218 tweets, with six polarity levels were split into two sets from the TASS challenge, were used [73]. Firstly, the tweets provided were shuffled and then the first set, hereafter the training set, was created with the first 6496 tweets (approximately 90% of the dataset), and the second set, hereafter the validation set, was composed of the rest 722 tweets (approximately 10% of the dataset). The training set was used to fit a Support Vector Machine (SVM) using a linear kernel with $C = 1$, weights inversely proportional to the class frequencies, and using the one-against-rest multiclass strategy. The validation set was used to select the best classifier using as performance the score $F1$- or F-measure. This measure considers both the precision and the recall. The $F1$-score can be interpreted as a weighted average of the precision and recall, where an $F1$-score reaches its best value at 1 and worst at 0.

The first step was to model the data using different transformations, namely Latent Dirichlet Allocation (LDA) using an online learning proposed by Hoffman in [74], Latent Semantic Indexing (LSI), and TF-IDF. **Figure 4** presents the score F1, in the validation set, of an SVM using either LSI or LDA with normalised text, different levels of q-gram (4 and 5 g), and the number of topics is varied from 10 to 500 as well. It is observed that LSI outperformed LDA in all the configurations tested.

http://dx.doi.org/10.5772/64355

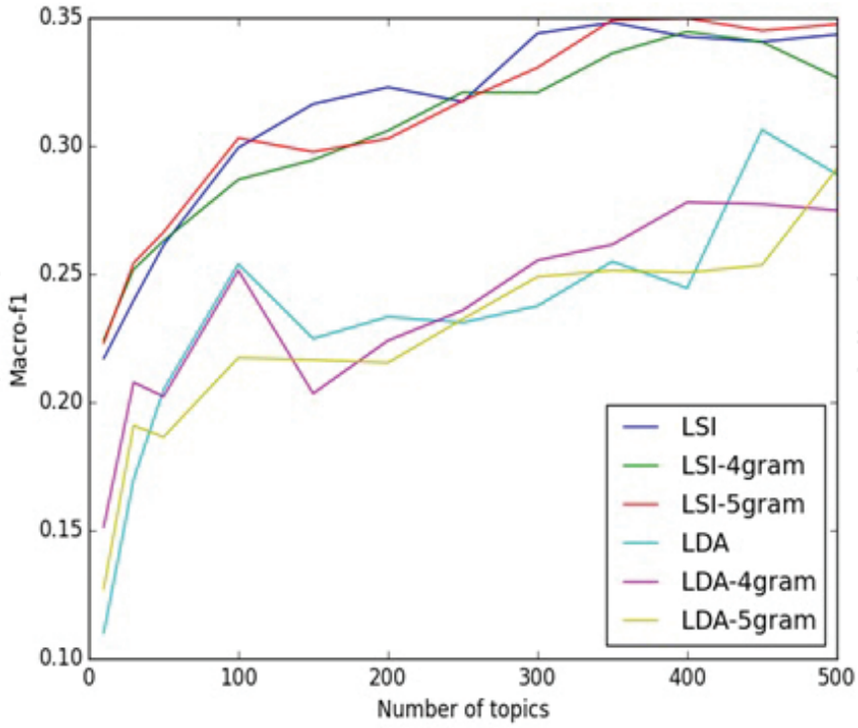

Figure 4. Performance of the various text transformations tested.

An equivalent performance was also observed when comparing the performance of normalised text, 4 and 5 g (**Figure 4**). Given that the implemented LSI depends on the order of the documents, more experiments are needed to know whether any particular configuration is statistically better than other. **Table 1** complements the information presented in **Figure 1**. **Table 1** presents the score F1 per polarity and the average (Macro-F1) for different configurations.

Table 1 is divided into five blocks, the first and second correspond to an SVM with LSI and TF-IDF, respectively. It is observed that TF-IDF outperformed LSI; within LSI and TF-IDF, it can be seen that 5 and 4 g got the best performance in LSI and TF-IDF, respectively. The third row block presents the performance when the features are a direct addition of LSI and TF-IDF; here, it is observed that the best performance is with 4 g. The fourth row block complements the previous results by presenting the best performance of LSI and TF-IDF, that is, LSI with 5 g and TF-IDF with 4 g. It is observed that this configuration has the best overall performance in P+, N, none and average (Macro-F1). Finally, the last row block gives an indication of whether the phonetic transformation is making any improvement. One major conclusion of this work is that the phonetic transformation is making a small difference.

As a final contribution, a set of experimental statistics were generated for the National Institute of Geography and Statistics (or INEGI from its Spanish name), yielding a map of subjective well-being for conterminous Mexico (**Figure 5**). This map reflects the importance of geospatial information, harvested from social media, because it allows us to measure subjective well-being on finer spatial and temporal resolutions than traditional methods.

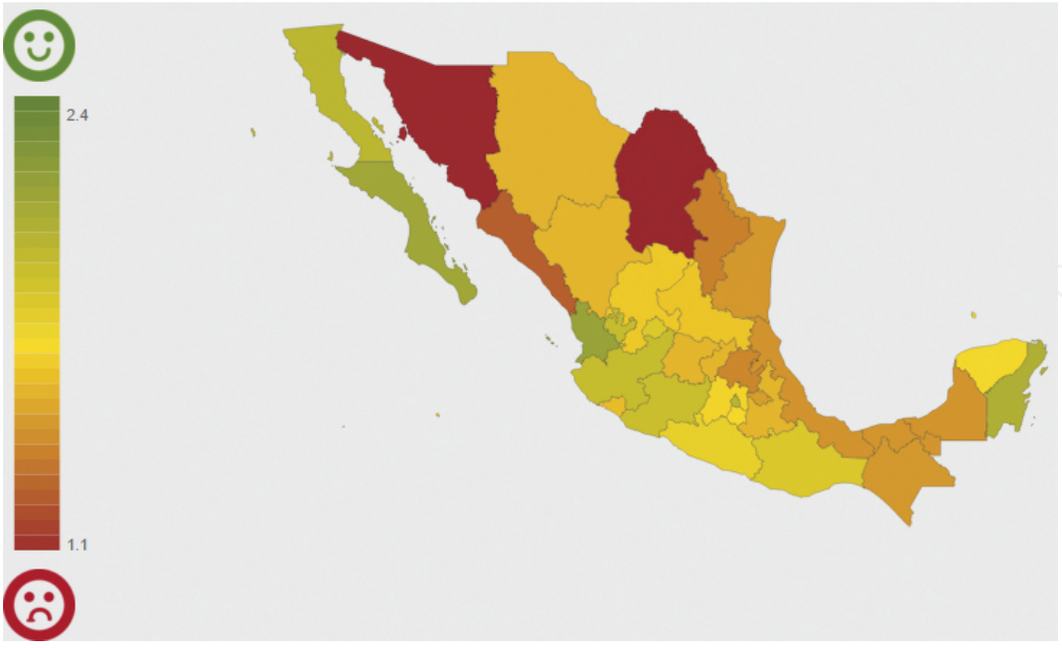

Figure 5. Subjective well-being map of Mexico based on the sentiment analysis of tweet messages.

3.2. Characterisation of migratory flow patterns of highly qualified people in Mexico

Many real systems—social, technological, biological and information—can be described as networks. We have only found few studies that treat migration from this perspective in the literature: one focusing on multiscale mobility in the United Kingdom [75], another dealing with internal migration in the United States [76], a global migration study stressing the flows between the OECD countries [77] and global flows [78, 79].

This case study treats the characterisation of migration flows of highly qualified human resources (defined by means of academic achievement—people with undergraduate degrees and those with graduate degrees—and people in knowledge-intensive occupations) in 59 Mexican metropolitan areas [80]. Data refer to the change of residence in the last 5 years, that is, recent migration was obtained from the 2010 General Population Census [81]. A common practice in migration studies is to aggregate data according to the analysis unit. In this case, starting with the origin-destination matrix, networks are built and then characterised. Furthermore, the square matrix is transformed into an array of minimum information that avoids redundancy and also allows for the dynamic exploration of flows between metropolitan areas.

Even though non-spatial visualisations reveal important properties of networks, it is interesting to try and shed some light on whether migration flows exhibit behaviour with strong geographical components.

Figure 6 shows the 'graduates' network. This network is partitioned in five communities and has the highest Q-value (0.66), implying a reasonable quality of the partition. It is worth noting

that the three largest metropolitan areas belong to different communities. Also, Mexico City encompasses almost half (23) of the metropolitan areas and its community is spread out throughout the whole country. By contrast, there is one community that consists of only one member and another one of only two members, both located in the centre of the country.

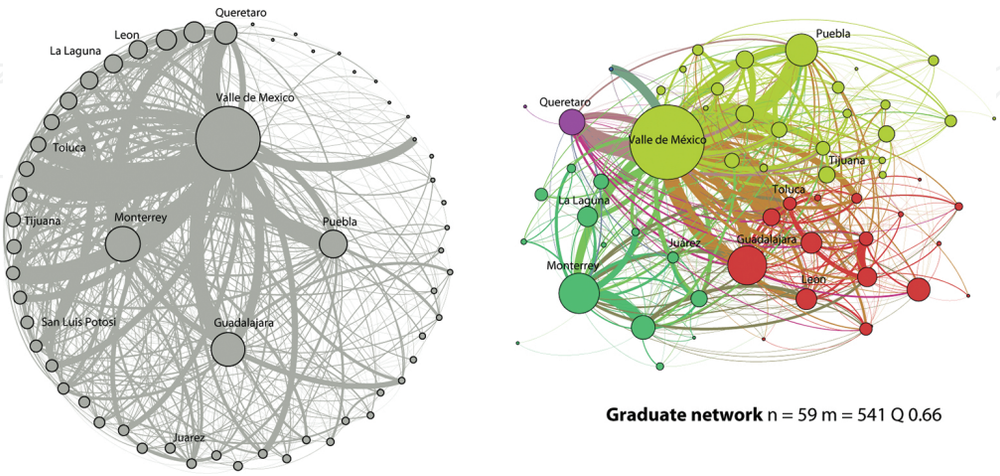

Figure 6. Graduates' migration network. Left: circular layout, showing labels for the 10 largest metropolitan areas; size is relative to the betweenness centrality parameter of the network. Right: nodes are coloured according to their community and the edges according to the source node.

An important characteristic of this study is network visualisation. By means of geographical visualisation, some network features can be highlighted according to node parameters. It also allows the identification of special structures in flow patterns.

Given the difficulty to explore flows and contextual elements related to the metropolitan areas, two separate interactive visualisations were prepared for this case. One uses Tableau Public and contains the analysis for community and role detection [82]. It also contains contextual data for each metropolitan area. The second is a geographic visualisation with special filters and functionality to explore the flows.

Tableau allows seeing the geographical arrangement of communities and the roles each metropolitan area plays (**Figure 4**). For the more dense networks—'undergraduates' and 'knowledge-intensive occupations'—there is an evident geographical component: communities tend to group regionally. The 'graduates' network instead exhibits a much smaller geographical distance than its functional one. This trend has been verified in other studies of high-quality human resources migration [83, 84]. It is important to note that concentrated or disperse functional distances cannot be highlighted using conventional network visualisations.

The interactive edges were a custom-made solution using open-source software. The frontend was built with jQuery [85] and LeafletJS [86]. The intensity of the inward and outward flows for each metropolitan area is represented with different colours and the number of migrants with relative widths. This interactive tool allows the comparison between origins and destinations for the different groups considered. Clicking on a metropolitan area simplifies

available information in the visualisation by only showing flows corresponding to that metropolitan area (**Figures 7** and **8**).

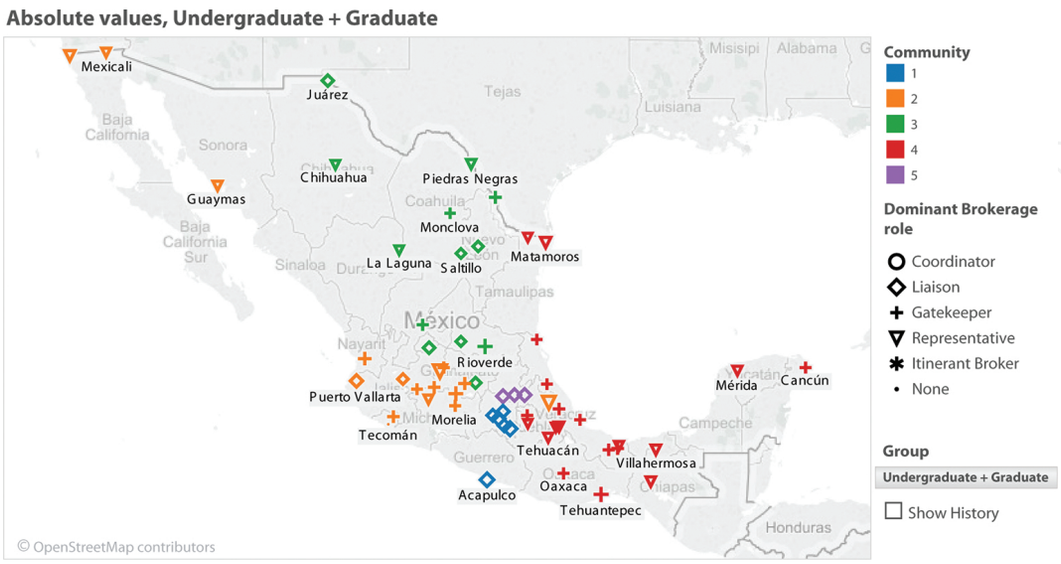

Figure 7. Communities for 'undergraduate' and 'graduate' migrations.

Figure 8. Flow visualisation for the metropolitan area of Cancun.

3.3. Volunteered geographic information for citizen empowerment

The case study presented in this section is set in a central neighbourhood in Mexico City: The Roma. The neighbourhood has experienced different stages throughout the years. At the

http://dx.doi.org/10.5772/64355

beginning of the twentieth century, it was considered to be high-class, rich people settled in the areas and several businesses experienced a florescence for several years. After a massive earthquake hit the city in 1985, many fled and the neighbourhood was partially abandoned for quite some time. Eventually, people who had lost their homes started to settle again in the neighbourhood, but by then it was not considered to be high class anymore. However, much of the architecture of the mid-1950s still remains even though many of these buildings have been occupied or have been used for different purposes other than residential. During the last decade, a gentrification process has been occurring in the neighbourhood, provoking poor people to be gradually expelled and richer people coming in. Because of the strong drastic changes that have occurred in it, the citizenry has started to notice many situations they consider to be harmful for their local environment. As a reaction, they have organised themselves and established an effective and fluent communication channel with their local authorities. After realising that they represent only a small portion of their municipality, they deemed it reasonable to explore the capabilities that crowdsourcing, VGI and participatory cartography could provide them.

For this, workshops were set up in order to find out about their needs and ideas. In an iterative process, the citizen part together with the scientific counterpart from CentroGeo converged on a list of variables to be collected on the field. This list represented the most pressing issues they could tackle for the moment and that were expected to be well received by the authorities in order to act and help ameliorate their situation. A list of six categories with several categories was agreed. A digital geospatial platform suitable for data collection on the web was set up. Due to time and budget constraints, it was not possible to provide them with native mobile apps. This platform consisted of purely free and open-source software: PostgreSQL/PostGIS [87, 88] for the backend, Bootstrap [89], jQuery and LeafletJS for the frontend and PHP [90] for the communication between both parts.

Citizens were in charge of data collection and quality assurance. The platform has the possibility to quickly get an idea about the spatial distribution of issues on the neighbourhood by means of a typical clustering strategy of collected data points. This is a very useful way for citizens to get an overall impression of what situations are persistent and, most importantly, where. Additionally, it is possible to create heat maps on the fly for the selected variables. This is useful for citizens to explore the possible existence of spatial correlation in the data they collected for different variables in their neighbourhood (**Figure 9**).

Overall, the case study was very successful in terms of allowing citizens to get more involved in noticing more details about everyday situations they face. It also helped them define possible courses of action to improve those situations in the neighbourhood. As of now, citizens are analysing all of the information they collected and establishing a plan to negotiate with their authorities. The process has helped them become more empowered because now they find themselves with data they did not think was possible to obtain. They thought they had to solely rely on what their local authorities could provide them and they have also found how they can come together for a greater good.

Also, it is worth mentioning that the maps that were obtained have been extremely useful to show where things are happening. This has been very helpful in increasing the citizenry's spatial awareness of their neighbourhood.

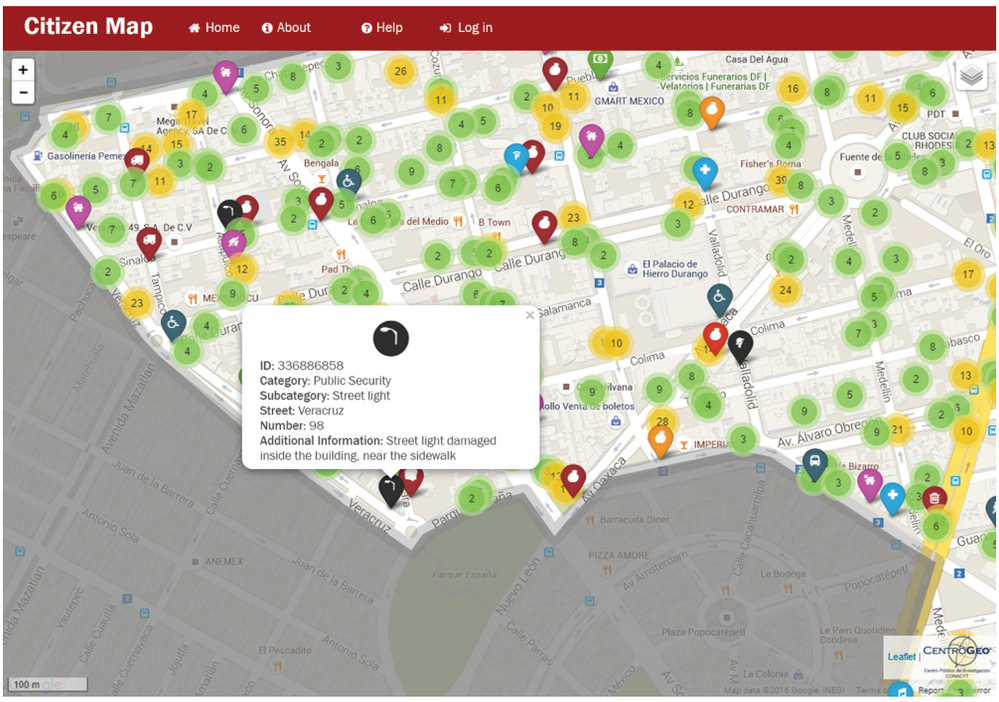

Figure 9. Citizen-mapping platform for the Roma neighbourhood showing clusters and categories.

3.4. Crime data analysis to support public safety in Mexico City

CentroGeo participated in the development of a geointelligence platform for Mexico City's Public Safety Ministry [91]. Back in 2004, this institution started georeferencing crime reports; in 2010, they already had enough experience in this task, but analytical capabilities were still short in order to extract useful information for decision-making. In this section, we present the implementation of a crime hotspot detection method that uses a spatio-temporal interaction graph.

The method mentioned in Section 2.4.2 was implemented in the context of Compstat-style planning and decision-making meetings that took place every week. A team of analysts would prepare comparative statistics and maps to establish police operations to focalised problems. Due to resource scarcity, it is imperative for public safety tasks to be prioritised. Hotspot detection for specific crime types was a first relevant criterion for decision-making.

As mentioned before, a first part of the process in mapping spatio-temporal hotspots consisted in the calculation of Knox's index together with the creation of the spatio-temporal interaction graph. Afterwards, the graph was characterised to identify the largest connected components, corresponding to priority areas (**Figure 1**).

http://dx.doi.org/10.5772/64355

Once these priority areas had been identified, human and material resources available to attack the problem were mapped. According to the detailed temporal patterns of incidents, it was possible to establish priority schedule tables for operating surveillance cameras in Mexico City (**Figure 10**).

Figure 10. Tactical planning map for the crime analysis study showing a hot area for larceny theft in Mexico City.

Implementing a geointelligence process in Mexico City's Public Safety Ministry was influenced both by the concept of geointelligence and by the institutional will to introduce a more fitting policing model for public safety in Mexico City. However, this has not been a linear process; instead, it has proven to be a complex, changing process entangling research and technical development results with daily demands emerging from the dynamics of the police institution.

3.5. Use of 3D vegetation modelling for forest inventorying Mexico City's Conservation Land

We present a case study of semiautomatic 3D forest generation through airborne laser scanner data over the Mexico City's Conservation Land (MCCL). Located in the southern fringe of Mexico City, the MCCL delivers important environmental services such as carbon sequestration, oxygen production, catchment, human recreation, among others, to the inhabitants of the city. However, its permanence has been threatened by urban sprawl during the past three decades generating several problems such as clandestine logging, illegal settlements and pollution [92]. The continuous monitoring and inventoring of this forested area will help authorities to preserve and improve this area. In this study, a 3D scene for an area of around 50 m^2 was generated using ALS data. Since the generation procedure and the accuracy assessment have been reported elsewhere [93, 94], here we only highlight the major processing steps and provide some theoretical insights of the 3D models.

3.5.1. ALS data processing

Point clouds acquired with the ALS50-II sensor flown by INEGI between November and December 2007 over the entire Basin of Mexico were employed in this study.

Basic processing prior to modelling surfaces with ALS data is the ground filtering and segmentation of the point cloud. The former refers to the segregation of ground points from the entire point cloud. Since feature heights are measured with respect to the ground, a bare-terrain surface must be first generated through interpolation of ground points. Then off-ground feature heights are normalised by subtracting the terrain elevation from the point cloud, and detection of objects of interest is conducted on the terrain-normalised dataset. For tree canopy detection, a fruitful approach is the watershed segmentation algorithm of the normalised digital height model with reversed z-coordinate. The segmentation procedure delineates watersheds that correspond, approximately, to tree crowns. Then, the segmentation is simply transferred to the points for the purpose of point selection.

3.5.2. Tree crown modelling

Points of individual trees were automatically selected using the segmentation information and best fit models were selected for each segment. A library of crown models was constructed from a generic revolution model of the form of Eq. (3), where (x, y, z) denotes a generic 3D point, and (u, θ) are the independent variables in the ranges [0,1] and $[0,\pi)$, respectively,

$$\begin{aligned} x &= C(u)\, r \cos\theta \\ y &= C(u)\, r \sin\theta \\ z &= h + (h-b)\, S(u) \end{aligned} \tag{3}$$

In this model, the crown size is represented by tree parameters, namely the maximum crown radius (r), the bottom crown height (b), and the top crown height (h), whereas the shape is represented by functions $C(u)$ and $S(u)$ defined in Eq. (4), where $c_1,\ldots,c_7$ denote the shape parameters

$$\begin{aligned} C(u) &= \left(c_1 - \left(c_2 + c_3 S(u)\right)^{c_4}\right)^{\frac{1}{c_5}} \\ S(u) &= 1 - (1 - u^{c_6})^{c_7} \end{aligned} \tag{4}$$

In order to simplify the model selection procedure, we computed the structure and location parameters from point statistics, and optimised shape parameters through a simplified least-squares orthogonal distance-fitting procedure. The orthogonal distance was computed only for a limited set of shape parameter combinations as given in **Table 2**, and then the least orthogonal distance model was selected as the best fit model. For visualisation purposes, the tree trunk was modelled as a cylinder of radius $0.1r$ and height b (**Figure 11**).

http://dx.doi.org/10.5772/64355

Model name	c_1	c_2	c_3*	c_4	c_5	c_6	c_7
Ellipsoid	1	1	2	2	2	2	2.4
Cylinder	2	0	1	0	2	1	1
Paraboloid	0	0	1	1	2	1	2
Hyperboloid	−1	i	−0.4142i	2	2	0.95	1.95
Cone	0	0	−i	2	2	1	1
Zparaboloid	0	0	−i	2	1	1	0.5

Table 2. Parameter combinations used for the crown shape model Eq. (4).

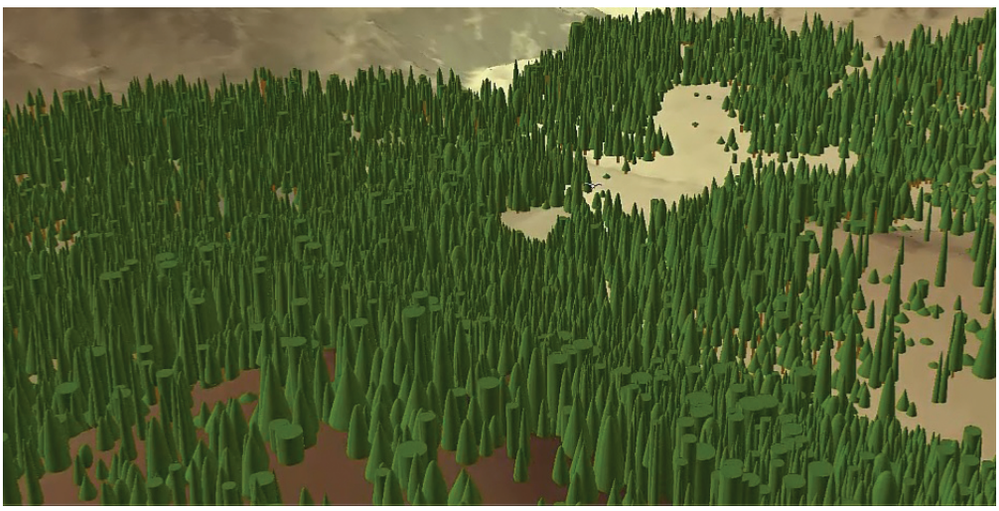

Figure 11. 3D visualisation of modelled forest from ALS data in the Mexico City Conservation Land.

The assessment of this product with ground truth data has shown the potential of ALS [93], especially for species communities exhibiting sparse distribution (such as *Pinus hartwegii* sp.), since limitations due to occlusion problems along dense species communities (such as *Abies religiosa* sp.) have also been reported suggesting the need to incorporate complementary TLS acquisitions. In any case, the utility of these techniques for large-scale inventorying is yet to be seen.

4. Concluding remarks

Geographic data collection has experienced a paradigm shift in terms of users being not only consumers but also generators. Traditionally, government agencies were in charge of collecting relevant information for different uses: cadastral, population and business censuses, vehicle registrars, natural resources, etc. However, it has become increasingly popular to be able to generate geographical data that do not necessarily adhere to governmental standards. Furthermore, it has become trending not only to collect but also to share these data in what

constitutes one of the pillars of neogeography: 'sharing location information with friends and visitors, help shape context, and conveying understanding through knowledge of place' [95], especially with all the mapping technologies available on the web [96].

This qualitative shift in the quantity and diversity of data that are gathered and examined has come with a shift in the techniques and technologies used to process and analyse information. In a seminal paper, Breiman talked about two cultures in data analysis: the 'classical' one, where data are modelled around a theoretical statistical distribution which, implicitly, assumes the kinds of processes producing the observations, and inferences are drawn from the distribution properties; and the 'algorithmic' one, where the focus is on extracting meaningful patterns and insights through the use of algorithmic models, without any assumptions about the mechanisms producing the observations [97]. This latter culture, stemming from the more empirical or applied fields, such as market research, electoral polling or computational biology, has gained momentum as the data we gather and analyse come, more often than not, from sources we cannot control (in a statistical sense), such as news outlets or social media feeds.

Author details

Jose Luis Silván-Cárdenas*, Rodrigo Tapia-McClung, Camilo Caudillo-Cos,
Pablo López-Ramírez, Oscar Sanchez-Sórdia and Daniela Moctezuma-Ochoa

*Address all correspondence to: jlsilvan@centrogeo.org.mx

Geography and Geomatics Research Centre – CentroGeo, Mexico City, Mexico

References

[1] Paradis, M. De l'arpentage à la géomatique (From surveying to geomatics). Can Surveyor. 1981;35(3):262–268.

[2] Kemp, K. Encyclopedia of Geographic Information Science. Sage; LA, USA 2008.

[3] Lillesand, TM, Kiefer, RW, Chipman, JW. Remote Sensing and Image Interpretation. 6th ed. USA: John Wiley & Sons, Inc.; NJ, USA 2008. 756p.

[4] McCaffrey, KJW, Jones, RR, Holdsworth, RE, Wilson, RW, Clegg, P, Imber, J, Holliman, N, Trinks, I. Unlocking the spatial dimension: digital technologies and the future of geoscience fieldwork. J Geol Soc. 2005;162(6):927–938.

[5] University College London. GeoKey [Internet]. 2015 [Updated: Nov 9 2015]. Available from: http://geokey.org.uk/. [Accessed: Jan 15 2016].

[6] Atzmanstorfer K, Resl R, Eitzinger A, Izurieta X. The GeoCitizen-approach: community-based spatial planning – an Ecuadorian case study. Cartogr Geogr Inf Sci. 2014;41(3): 1–12.

[7] Usahidi. Usahidi [Internet]. 2015 [Updated: Sep 15 2015]. Available from: https://www.ushahidi.com/

[8] Zook M, Graham M, Shelton T, Gorman S. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Med Health Policy. 2010;2(2):6–32.

[9] OpenTreeMap. OpenTreeMap [Internet]. 2015. Available from: https://www.opentreemap.org/. [Accessed: Nov 9 2015]

[10] California Academy of Sciences. iNaturalist [Internet]. 2015. Available from: http://www.inaturalist.org/. [Accessed: Nov 9 2015]

[11] Waze. Waze Live Map [Internet]. 2015. Available from: https://www.waze.com/livemap. [Accessed: Sep 9 2015]

[12] Sony Computer Science Laboratory Paris. Software Languages Lab VUB. NoiseTube [Internet]. 2015. Available from: http://www.noisetube.net/. [Accessed: Feb 17 2015]

[13] Crooks A, Pfoser D, Jenkins A, Croitoru A, Stefanidis A, Smith D, et al. Crowdsourcing urban form and function. Int J Geogr Inf Sci. 2015;29(5):720–741.

[14] Sui D, Goodchild M, Elwood S. Volunteered geographic information, the exaflood, and the growing digital divide. In: MF Goodchild, editor. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. Netherlands: Springer Science & Business Media; Netherlands 2013. pp. 1–12.

[15] Harvey F. To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. In: MF Goodchild, editor. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. Springer Netherlands; Netherlands 2013. pp. 31–42.

[16] Surowiecki J. The Wisdom of Crowds: Why the Many are Smarter Than the Few and How Collective Wisdom Shapes Business Economies Societies and Nations. New York: Doubleday; 2004. 296p.

[17] Spielman SE. Spatial collective intelligence? credibility, accuracy, and volunteered geographic information. Cartogr Geogr Inf Sci. Taylor & Francis; 2014;41(2):115–124.

[18] OpenSignal. Mexico's cell phone coverage [Internet]. 2015 [cited 2016 Feb 4]. Available from: http://opensignal.com/?lat=23.2897&lng=-101.8675&initZoom=6&isHeatMap=1

[19] Ball, GH, Hall, DJ, Stanford Research Institute, United States, Office of Naval Research, Information Sciences Branch. Isodata, a novel method of data analysis and pattern classification. Stanford Research Institute, Menlo Park, Calif, 1965.

[20] Duque, JC, Ramos, R, Suriñach, J. Supervised regionalization methods: a survey. Int Reg Sci Rev. 2007; 30:195–220. doi:10.1177/0160017607301605

[21] Kulldorff, M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Stat Soc Ser A Stat Soc. 2001; 164:61–72. doi:10.1111/1467-985X.00186

[22] Chen, Y-K, Wang, J-F. Segmentation of single- or multiple-touching handwritten numeral string using background and foreground analysis. IEEE Trans Pattern Anal Mach Intell. 2000; 22:1304–1317. doi:10.1109/34.888715

[23] Chuang, K-S, Tzeng, H-L, Chen, S, Wu, J, Chen, T-J. Fuzzy c-means clustering with spatial information for image segmentation. Comput Med Imaging Graph. 2006;30:9–15. doi:10.1016/j.compmedimag.2005.10.001

[24] Ng, HP, Ong, SH, Foong, KWC, Goh, PS, Nowinski, WL. Medical image segmentation using K-means clustering and improved watershed algorithm. In: 2006 IEEE Southwest Symposium on Image Analysis and Interpretation. Presented at the 2006 IEEE Southwest Symposium on Image Analysis and Interpretation; 2006. pp. 61–65. doi: 10.1109/SSIAI.2006.1633722

[25] Wang, S, Anselin, L, Bhaduri, B, Crosby, C, Goodchild, MF, Liu, Y, Nyerges, TL. CyberGIS software: a synthetic review and integration roadmap. Int J Geogr Inf Sci. 2013;27:2122–2145. doi:10.1080/13658816.2013.776049

[26] Huang, R, Yang, Q, Pei, J, Gama, J, Meng, X, Li, X (Eds.). Advanced Data Mining and Applications, Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009.

[27] Miller, HJ, Han, J (Eds.). Geographic Data Mining and Knowledge Discovery, 2nd ed. Boca Raton, FL: CRC Press; 2009.

[28] Frias-Martinez, V, Frias-Martinez, E. Spectral clustering for sensing urban land use using Twitter activity. Eng Appl Artif Intell. 2014;35:237–245. doi: 10.1016/j.engappai.2014.06.019

[29] Lee, R, Wakamiya, S, Sumiya, K. Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web. 2011;14:321–349. doi: 10.1007/s11280-011-0120-x

[30] Gabrielli, L, Rinzivillo, S, Ronzano, F, Villatoro, D. From tweets to semantic trajectories: mining anomalous urban mobility patterns. In: J Nin, D Villatoro, editors. Citizen in Sensor Networks, Lecture Notes in Computer Science. Springer International Publishing; Switzerland 2014. pp. 26–35.

[31] Boettcher, A, Lee, D. EventRadar: a real-time local event detection scheme using twitter stream. IEEE. 2012;358–367. doi:10.1109/GreenCom.2012.59

[32] Kim, T, Huerta-Canepa, G, Park, J, Hyun, SJ, 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

International Conference on Social Computing (SocialCom). Presented at the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom); 2011. pp. 806–809. doi: 10.1109/PASSAT/SocialCom.2011.135

[33] Kim, T., Huerta-Canepa, G., Park, J., Hyun, S.J., Lee, D., 2011. 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). Presented at the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), IEEE, Boston, MA, pp. 806–809. doi:10.1109/PASSAT/SocialCom.2011.135

[34] Padmanabhan, A, Wang, S, Cao, G, Hwang, M, Zhang, Z, Gao, Y, Soltani, K, Liu, Y. FluMapper: a cyberGIS application for interactive analysis of massive location-based social media. Concurr Comput Pract Exp. 2014;26:2253–2265. doi:10.1002/cpe.3287

[35] Wang, S, Hu, H, Lin, T, Liu, Y, Padmanabhan, A, Soltani, K. CyberGIS for data-intensive knowledge discovery. SIGSPATIAL Spec. 2015;6:26–33.

[36] Wang, Y, Weinacker, H, Koch, B. A lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modeling in forest. Sensors. 2008;8. 3938-3951

[37] Li, L., Xi, Y., 2011. Research on Clustering Algorithm and Its Parallelization Strategy, in: 2011 International Conference on Computational and Information Sciences (ICCIS). Presented at the 2011 International Conference on Computational and Information Sciences (ICCIS), IEEE, Chengdu, China, pp. 325–328. doi:10.1109/ICCIS.2011.223

[38] Marozzo, F, Talia, D, Trunfio, P. P2P-MapReduce: parallel data processing in dynamic cloud environments.., J Comput Syst Sci. JCSS Special Issue: Cloud Computing. 2012;78:1382–1402. doi: 10.1016/j.jcss.2011.12.021

[39] Nystuen, JD, Dacey, MF. A graph theory interpretation of nodal regions. Pap Reg Sci. 2005;7(1):29–42.

[40] Graizbord, B. Geografía del transporte en el area metropolitana de la Ciudad de México. Mexico, AC: El Colegio de; 2008. 386p.

[41] Súarez-Lastra, M, Delgado-Campos, J. Urban structure and efficiency. Job accessibility, residential location and income to Mexico City's metropolitan area. Econ Soc y Territ. 33 2007;6(23):693–724.

[42] Bender-deMoll, S. Potential human rights uses of network analysis and mapping. A report to the Science and Human Rights Program of the American Association for the Advancement of Science. 2008. Available at http://skyeome.net/wordpress/wp-content/uploads/2008/05/Net_Mapping_Report.pdf Accessed Feb 12 2016.

[43] Fortunato, S. Community detection in graphs. Phys Rep 2008;486(3–5):75–174.

[44] Blondel, VD, Guillaume, J-L, Lambiotte, R, Lefebvre, E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008;2008(10):P10008.

[45] MacEachren, AM, Gahegan, M, Pike, W, Brewer, I, Cai, G, Lengerich, E, et al. Geovisualisation for knowledge construction and decision support. IEEE Comput Graph Appl. 2004;24(1):13–17.

[46] Andrienko, G, Andrienko, N, Jankowski, P, Keim, D, Kraak, M-J, MacEachren, A, et al. Geovisual analytics for spatial decision support: Setting the research agenda. Int J Geogr Inf Sci. 2007;21(8):839–857.

[47] Smith, S, Bruce, C. CrimeStat III. Workbook. Washington, D.C.: National Institute of Justice; Washington, DC, USA 2008. 113p.

[48] Knox, EG, Bartlett, MS. The detection of space-time interactions. J R Stat Soc Ser C (Appl Stat). 1964;13(1):25–30.

[49] Jacquez, GM, Greiling, DA, Durbeck, H, Estberg, L, Do, E, Long, A, Rommel B. ClusterSeer User Guide 2: Software for identifying disease clusters. Ann Arbor, MI: TerraSeer Press 2002.

[50] Pettit, C, Widjaja, I, Russo, P, Sinnott, R, Stimson, R, Tomko, M. Visualisation support for exploring urban space and place. ISPRS Ann Photogramm Remote Sens Spat Inf Sci 2012;I–2(1):153–158.

[51] Google. Google Geo Developers Blog [Internet]. 2012 [cited 2016 Feb 15]. Available from: http://googlegeodevelopers.blogspot.mx/2012/06/powerful-data-visualisation-with.html

[52] Xiao, N, Chun, Y. Visualizing migration flows using kriskograms. Cartogr Geogr Inf Sci. 2009;36(2):183–191.

[53] Boyandin, I, Bertini, E, Bak, P, Lalanne, D. Flowstrates: an approach for visual exploration of temporal origin-destination data. Comput Graph Forum. 2011;30(3):971–980.

[54] Tobler, W. Experiments in migration mapping by computer. Am Cartogr. 1987;14(2): 155–163.

[55] ESRI. ArcGIS 3D Analyst [Internet]. 2011 [Updated: 2016]. Available from: http://www.esri.com/software/arcgis/extensions/3danalyst/. [Accessed: 12/03/2016]

[56] ESRI. Esri CityEngine [Internet]. 2011 [Updated: 2016]. Available from: http://www.esri.com/software/cityengine. [Accessed: 12/03/2016]

[57] Morsdorf, F, Meier, E, Kötz, B, Itten, KI, Dobbertin, M, Allgöwer, B. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. Remote Sens Environ. 2004;92:353–362.

[58] Kato, A, Moskal, LM, Schiess, P, Swanson, ME, Calhoun, D, Stuetzle, W. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sens Environ. 2009;113:1148–1162.

[59] Rutzinger, M, Pratihast, AK, Oude Elberink, S, Vosselman, G. Detection and modelling of 3D trees from mobile laser scanning data. Int Arch Photogramm Remote Sens Spat Inf Sci. 2010;38:520–525.

[60] Sampath, A, Shan, J. Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds. IEEE Trans Geosci Remote Sens. 2010;48(3):1554–1567.

[61] Faugeras, O, Robert, L, Laveau, S, Csurka, G, Zeller, C, Gauclin, C, Zoghlami, I. 3-d reconstruction of urban scenes from image sequences. Comput Vis Image Understand. 1998;69(3):292–309.

[62] Verbree, E, Maren, GV, Germs, R, Jansen, F, Kraak, M-J. Interaction in virtual world views-linking 3D GIS with VR. Int J Geogr Inf Sci. 1999;13(4):385–396.

[63] Goodchild, M. Time, space, and GIS. Past Place: Newsl Hist Geogr Spec Group Assoc Am Geogr. 2006;14(2):8–9.

[64] Langran G, Chrisman NR. A framework for temporal geographic information. Cartogr Int J Geogr Inf Geovisualisation. 1988;25(3):1–14.

[65] Hagerstrand T. Innovation diffusion as a spatial process. Chicago, USA: Univerity of Chicago Press; 1968.

[66] Hagerstrand T. Diorama, path and project. Tijdschrift voor economische en sociale geografie. Swedish 1982;73(6):323–339.

[67] Peng, T, Zuo, W, He, F. SVM based adaptive learning method for text classification from positive and unlabeled documents. Knowl Inf Syst. 2008;16(3):281–301.

[68] Pang, B, Lee, L. Opinion mining and sentiment analysis. Found Trends Inf Retr. 2008;2(1–2):1–135.

[69] da Silva, NFF, Hruschka, ER, Hruschka, ER. Tweet sentiment analysis with classifier ensembles. Decis Supp Syst. 2014;66:170–179.

[70] Shahbaz, M, Guergachi, A. Sentiment miner: a prototype for sentiment analysis of unstructured data and text. In: 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE); 2014; Canada. IEEE; pp. 1–7.

[71] Liu, SM, Chen, J-H. A multi-label classification based approach for sentiment classification. Expert Syst Appl. 2015;42(3):1083–1093.

[72] Padró, L, Stanilovsky, E. Freeling 3.0: towards wider multilinguality. In: Nicoletta C. et al., editors. Eighth International Conference on Language Resources and Evaluation; Istanbul, Turkey; 2012.

[73] Martínez-Cámara, E, Martín-Valdivia, MT, Urena-López, LA, Montejo-Ráez, AR. Sentiment analysis in twitter. Nat Lang Eng. 2014;20(1):1–28.

[74] Hoffman, M, Bach, FR, Blei, DM. Online learning for latent Dirichlet allocation. In: Lafferty JD, Williams CKI, Shawe-Taylor J, Zemel RS, Culotta A, editors. Advances in Neural Information Processing Systems; Curran Associates, Inc.; NY, USA 2010. pp. 856–864.

[75] Askar, D, House, T. Complex patterns of multiscale human mobility in United Kingdom methodology data. Warwick, UK; 2010. Accessed February 12 2015. Available at https://www2.warwick.ac.uk/fac/cross_fac/complexity/study/emmcs/outcomes/studentprojects/askar_m2.pdf

[76] Maier, G, Vyborny, M. Internal migration between US States – a social network analysis. European Regional Science Association; Vienna, Austria 2005.

[77] Tranos, E, Gheasi, M, Nijkamp, P. International migration: a global complex network. Environ Plan B Plan Des. 2015;42(1):4–22.

[78] Fagiolo, G, Mastrorillo, M. International migration network: topology and modeling. Phys Rev E Stat Nonlin Soft Matter Phys. 2013; 88(1):012812.

[79] Porat, I, Benguigui, L. World migration degree. Global migration flows in directed networks. 2015. Available from http://arxiv.org/abs/1511.05338. Accessed February 8 2016.

[80] CONAPO (Consejo Nacional de Población), INEGI (Instituto Nacional de Estadística y Geografía). Delimitation of Mexico's metropolitan areas 2010. Mexico, 2013.

[81] INEGI (Instituto Nacional de Estadílstica y Geografía). Censo de Población y Vivienda 2010: Interactive data query [Internet]. 2011 [cited 2015 Feb 1]. Available from: http://www.inegi.org.mx/lib/olap/consulta/general_ver4/MDXQueryDatos.asp?#Regreso&c=27770

[82] Tableau Foudation. Tableau [Internet]. 2003 [Updated: 2016]. Available from: http://www.tableau.com/. [Accessed: Nov 2015]

[83] Zelinsky, W. The hypothesis of the mobility transition. Geogr Rev. 1971;61(2):219–249.

[84] Abel, JR, Deitz, R. Do colleges and universities increase their region's human capital? J Econ Geogr. 2012;12(August 2011):667–691.

[85] jQuery Foundation. jQuery [Internet]. 2016 [Updated: 2016]. Available from: https://jquery.com/. [Accessed: Jan 2016]

[86] Vladimir Agafonkin. Leaflet [Internet]. 2015 [Updated: 2016]. Available from: http://leafletjs.com/. [Accessed: Feb 2016]

[87] PosgreSQL. PosgreSQL [Internet]. 2009 [Updated: 2013]. Available from: http://www.postgresql.org.es/. [Accessed: Jan 2016]

[88] OSGeo. PostGIS [Internet]. 2016 [Updated: 2016]. Available from: http://postgis.net/. [Accessed: Feb 2016]

[89] @mdo and @fat. BootStrap [Internet]. 2015 [Updated: 2016]. Available from: http://getbootstrap.com. [Accessed: Jan 2015]

[90] The PHP Group. PHP [Internet]. 2001 [Updated: 2016]. Available from: http://php.net. [Accessed: Jan 2016]

[91] Martínez-Viveros, E, Chapela, J, Morales-Gamas, A, Caudillo-Cos, C, Tapia-McClung, R, Ledesma, M, et al. Construction of a web-based crime geointelligence platform for Mexico city's public safety. In: Leitner M, editor. Crime Modeling and Mapping Using Geospatial Technologies SE-18. Springer Netherlands; Netherlands 2013. pp. 415–439.

[92] Aguilar, GA, Ward, MP. Globalization, regional development, and mega-city expansion in Latin America: analyzing Mexico City´s peri-urban hinterland. Cities. 2003;20(1):3–21.

[93] Silván-Cárdenas, JL. A segmentation method for tree crown detection and modelling from LiDAR measurements. Pattern Recognition. LNCS. 2012;7329:65–74.

[94] Silván-Cárdenas, JL, Corona-Romero, N, Galeana-Pizaña, JM, Nuñez-Hernández, JM, Madrigal Gómez, JM. Geospatial technologies to support coniferous forests research and conservation efforts in Mexico. In: Weber RP, editor. Old-Growth Forests and Coniferous Forests: Ecology, Habitat and Conservation. New York: Nova Science Publishers; pp. 67–123.

[95] Turner, A. Introduction to neogeography. O'Reilly Media, Inc.; CA, USA 2006. 54p.

[96] Haklay, M, Singleton, A, Parker, C. Web mapping 2.0: the neogeography of the GeoWeb. Geogr Compass. 2008;2(6):2011–2039.

[97] Breiman, L. Statistical modeling: the two cultures. Stat Sci. 2001;16:199–215.

IntechOpen