# Definition of Strategies for Crime Prevention and Combat Using Fuzzy Clustering and Formal Concept Analysis

**Article** *in* International Journal of Uncertainty Fuzziness and Knowledge-Based Systems · June 2018

DOI: 10.1142/S0218488518500216



**4 authors**, including:

Marcos Evandro Cintra
Universidade Federal Rural do Semi-Árido - UFERSA



Angélica F. Castro
Universidade Federal Rural do Semi-Árido - UFERSA

International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems


# Definition of strategies for crime prevention and combat using fuzzy clustering and formal concept analysis

Adriana M. Guimaraes*, Marcos E. Cintra†, Angelica C. Felix*, Danniel L. Cavalcante*

*Centre of Exact and Natural Sciences, Federal Rural University of the Semi Arid, Brazil**

*Institute of Science and Technology, Federal University of Sao Paulo, Avenida Cesare Mansueto Giulio Lattes, 1201 - Eugenio de Mello, CEP: 12247-014, Sao Jose dos Campos, SP, Brazil*†

Received (received date)
Revised (revised date)

Public security has always been an important research topic. In this sense, machine learning algorithms have been used to extract knowledge from criminal databases, which usually maintain records in order to generate statistics. The automatic extraction of knowledge from such databases allows the improvement and planning of strategies to prevent and combat crimes. Accordingly, in this work different models related to public security are presented. Such models are based on clustering algorithms, on the analysis of formal concept techniques, and on the analysis of crime record data collected in the city of Mossoro, Brazil. The two types of models generated are: i) concept lattices with crime patterns; ii) criminal hot spot maps. We also produced a ranking of dangerousness for neighbourhoods of Mossoro. The Fuzzy K-Means clustering algorithm was used to obtain criminal hot spots, which indicate locations with high crime incidence. Formal concept analysis was used for extracting visual models describing patterns that characterize criminal activities. Such models have the form of conceptual lattices that provide graphical displays which can be used for defining strategies to combat and prevent crime. The models were first empirically evaluated and then analysed by public security experts, who provided positive feedback for their practical use. The advantages of the automatically generated models presented in this paper are many, including the short time to produce such models, the variety of different models that can be generated for specific regions and periods of days, months, or years, the graphical characteristic of such models that allow a fast analysis of them, as well as the use of large amounts of data, which are infeasible activities to be done by human experts.

*Keywords*: Public security; Crime prevention; Criminal hot spots; Fuzzy Clustering; Clustering algorithms; Formal concept analysis.

## 1. Introduction

Criminality can be defined as "the state of being illegal; quality, state, or fact of being illegal; a criminal practice or the act or series of acts that constitute a crime" [1].

*adrianamara.guimaraes@gmail.com,angelica@ufersa.edu.br,danniel@ufersa.edu.br
†mecintra@gmail.com

Criminality generates high economic losses, affecting the welfare of the population. The Brazilian government has been investing in public security, including in police intelligence departments in order to deal with the increasing number of crime incidents and reduced economical resources. Such investments is a world tendency due to the impact of crime on economy. For instance, Table 1 presents information about the costs incurred for Public Security in relation to the Gross Domestic Production (GDP) and rates of homicide for some selected countries for the year of 2014 [6]. Notice that 27 countries are included in the European Union data.

| Country | GDP % | Homicides | Rate of homicide |
|---|---|---|---|
| Brazil | 1.26 | 50,806 | 25.2 |
| Chile | 0.80 | 550 | 3.1 |
| European Union | 1.30 | 5,539 | 1.1 |
| France | 1.38 | 665 | 1.0 |
| Germany | 1.06 | 662 | 0.8 |
| Guatemala | 0.70 | 6025 | 39.9 |
| United Kingdom | 1.56 | 653 | 1.0 |
| USA | 1.02 | 14,827 | 4.7 |

Table 1. Cost incurred for public security in relation to GDP and rates of homicide.

For police activities, it is necessary to invest in vehicles, surveillance cameras and other types of equipments, as well as on the maintenance of human resources, implicating in high costs. However, if more investment on information and intelligence is made, it is known that such costs can be reduced [7].

### 1.1. *Related Work*

A throughout review of the literature on implementing hotspot policing can be found in [23]. Another relevant literature volume for the understanding of crime mapping can be found in [10].

An analysis of criminal indexes applying regression based on the Poisson distribution is described in [21]. The crime indexes were calculated using the size of the population. Thus, considering the index of crimes of a given place, the number of violations should be proportional to number the of inhabitants in that place.

Musdholifah, Hashim, and Wasito [19] analysed the grouping of criminal data with the application of an iterative algorithm known as Iterative Local Gaussian Clustering (ILGC). The ILGC algorithm combined the closest neighbours to the kernel density to define their groups, in order to iteratively obtain the best sites. The validation of the proposal was carried out by comparing its results to Self-Organizing Maps and the K-means algorithm.

In [14], the authors use support vector machine for the prediction of crime hot spots. In [5], on the other hand, the authors used genetic algorithms to define crime hot spots.

In reference [29], the authors used kernel density for detection of criminal hot spots from criminal records of residence theft. That study analysed a database with records dated between 1999 and 2007, and describes an extended method for exploring nonstationarity of local mechanisms of residential burglaries through a temporal variant of geographically weighted regression within a space-time cube. The proposed model which is aided with $3D$ volume rendering not only identifies variations of space-time clusters, but also visualizes the dynamics of criminal mechanisms specifically. The proposed method also considers and includes a space-time trade-off into its weighting procedure.

In [?], the authors propose to detect patterns of crimes in crime databases.

A new visualization method of crime patterns based on geographical data using FCA was proposed in [16]. The author considered the types of crimes analysed as the attributes and the city neighbourhoods as objects. Therefore, the identification patterns by the lattice allowed to define the crimes that were concentrated in certain neighbourhoods. Another FCA based proposal can be found in [15], which used a larger set of attributes to obtain better results. In addition to the type of the committed crimes, the author used the following attributes: sex, age and level of education.

### 1.2. *Contributions*

Mossoro is a city in the North-east region of Brazil with a population of 284,288 inhabitants [3]. Although Mossoro can be considered a fairly safe city for Brazilian patterns, in recent years the feeling of insecurity has grown among citizens. The Scientific and Technical Police Institute [a], responsible for the distribution of crime statistics has shown alarming numbers related to the increase in criminality in the city [26]. For instance, according to data from the Sangari Institute [13], homicide rates increased from 132 homicides in 2009 to 192 in 2014.

For this reason, the investment in public security is mandatory. In fact, as the financial resources to public security are limited, the use of Artificial Intelligence techniques represents a low cost option to create strategies to crime combat and prevention. This way, the obtained models with the data from Mossoro allows the definition of strategies for the planning of police activities and were positively received by the local Police Department. Moreover, such models can be easily updated and scaled with the use of new data and can be generated to distinct cities easily.

Based on the advantages of using the available crime data of Mossoro, as well as in other cities, we proposed the automatic extraction of two types of models to support police activities and the definition of dangerousness for neighbourhoods, as described next:

(1) The use of the Fuzzy K-Means clustering algorithm allowing the identification of sites with high crime concentration (criminal hot spots) based on the location

[a] `http://www.itep.rn.gov.br/`

data available. To the best of our knowledge, this is the first time Fuzzy K-Means is used for the definition of criminal hot spots);

(2) The use of Formal Concept Analysis (FCA) to process criminal data, allowing the extraction of crime patterns for specific periods of the day and types of crimes. Four types of basic crimes (theft, robbery, homicide, drug traffic) were analysed using the FCA theory to generate lattices according to the concentration of such crimes throughout specific periods of the day (morning, afternoon, evening, night);

(3) As a third contribution, we also produced a ranking of the neighbourhoods of Mossoro according to their level of dangerousness by analysing the collected data.

The remainder of this paper is organized as follows. In Section 2 we introduce the topics of criminal analysis and criminal hot spots and we discuss relevant concepts about clustering algorithms and Formal Concept Analysis. Section 3 describes the adopted methodology to collect and pre-process the data and to generate the beforementioned models. Section 4 shows the results, followed by the final considerations and future work in Section 5.

## 2. Theoretical Background

In this section, we introduce the topics of criminal analysis, criminal hot spots, clustering algorithms, and Formal Concept Analysis.

### 2.1. *Criminal Analysis*

Criminal analysis is defined as "the use of a collection of methods to plan actions and public security policies, to obtain, organize, analyse, and interpret data in order to make conclusions from such data" [22].

Criminal analysis can be performed using statistical methods of data processing to obtain useful information that can be used to understand the causes of criminality. Such task include two basic operations, crime pattern and crime trend assessment [8]. Criminal analysis provides support to operational and administrative areas of the Police during the planning and distribution of resources in order to prevent and suppress criminal actions.

In fact, crime patterns can be obtained by observing at least one feature in a set of events that repeats along time. Such characteristic can be the location, date, time, or the profile of a victim or offender, for instance. The analysis of crime trends allows the evaluation of crime types in a given location for a period of time, *i.e.*, whether a certain crime type is increasing, stabilizing or decreasing.

Machine learning techniques can be easily applied to generate models from collected data.

### 2.2. *Criminal Hot Spots*

Criminal Hot Spots can be defined as geographic areas with high concentration of criminal incidence, *i.e.*, an area in which the number of criminal events is considered significant. In other words, criminal hot spots are places where people have high risk of being victimized [7]. Such places that present high crime concentration are usually determined by the following characteristics: a) amount of events in the place; b) spatial information; c) time of crime; and d) type of event [30].

Criminal hot spots are usually presented as point maps [20]. Hot spot maps are popularly known as maps with pins which represent crimes. Such maps are usually displayed on walls of police stations. As a computational tool, hot spot maps can receive more information, such as codes to describe the type, date, and time of crimes. Figure 2.2 shows an example of a criminal hot spot map with information on homicides for the city of Houston, USA.

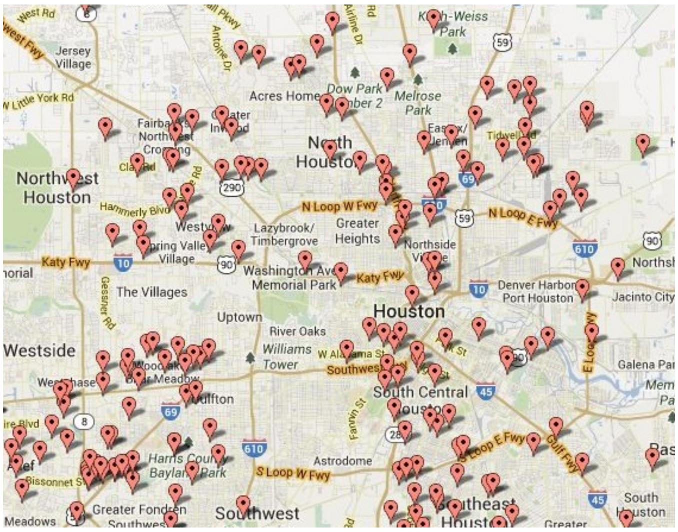


Figure 1. Criminal Hot Spot map of homicides in the city of Houston, USA.

The Houston hot spot map in Figure 2.2 shows pins in the places where homicide crimes occurred in 2014.

### 2.3. *Clustering Algorithms*

Clustering algorithms can be defined as "procedures to organize objects into groups whose members are similar in some way". A cluster is, therefore, a collection of objects which are "similar" among them and are "dissimilar" in relation to the objects belonging to other clusters [18].

Usually, clustering algorithms organize objects into groups by comparing them according to certain distances. The main distances used in clustering algorithms are: Euclidean, Manhattan and Mahalanobis [12].

An important question that needs to be addressed before applying clustering algorithms is determining the number of clusters that there should be in the data. Usually, this parameter is unknown and set empirically. In fact, there might be no definite value for the number of clusters in the data. However, it is possible to find proposals for the automatic estimation of the number of clusters in the literature [17].

K-Means [24] is one of the most well-known clustering algorithms. K-Means aims at finding the best division of $n$ entities in $k$ groups, so that the total distance between the members of each group to their corresponding centroid is minimized. Formally, the goal is to partition the $n$ entities into $k$ sets $S_i, i = 1, 2, ..., k$, in order to minimize the within-cluster sum of squares. K-Means is a greedy, computationally efficient technique, and one of the most popular clustering algorithms. For the fuzzy version of K-Means, Fuzzy K-Means, every point has a degree of pertinence to each cluster, as in the fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the centre of the cluster.

## 2.4. *Formal Concept Analysis*

Formal Concept Analysis (FCA) is a mathematical technique to extract characteristics and structures from data [27].The basic data structure in FCA is the formal context, which is normally displayed in a table form where the columns represent the attributes and the rows the objects (see Table 2). In other words, a formal context is a representation of the relation between objects and their attributes.

It is possible to generate a conceptual lattice from a formal context. A conceptual lattice is basically a graph whose vertices correspond to the formal concepts represented by sets of examples or attributes [4].

Tables 2 and 3 present a toy example showing information on the age, sex (M for male and F for female), and hair colour of 6 people.

| **Name** | **Age** | **Sex** | **Hair Colour** |
|---|---|---|---|
| Andy | 48 | M | Black |
| Linda | 29 | F | Black |
| Mark | 23 | M | Brown |
| Martina | 46 | F | Blond |
| Mike | 18 | M | Brown |
| Suzy | 17 | F | Blond |

Table 2. A toy example of an Attribute × Value table used with FCA.

As FCA only works with binary attributes, a scaling process, *i.e.*, the transformation of attributes into binary ones, is usually required. Table 3 shows the binary version of attributes Age, Sex and Hair Colour of Table 2.

| **Name** | **Age** | | | **Sex** | | **Hair Colour** | | |
|---|---|---|---|---|---|---|---|---|
| | $\leq$ ***20*** | $[20, 30)$ | $>$***30*** | ***M*** | ***F*** | ***Blond*** | ***Brown*** | ***Black*** |
| Andy | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| Linda | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| Mark | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| Martina | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| Mike | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| Suzy | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |

Table 3. Formal Context based on Table 2.

In Table 3, attribute Age was discretized (or scaled), generating three binary attributes. The multi-valued attributes Sex and Hair Colour generated a binary attribute for each of their values. Notice that FCA does not require a class attribute. Figure 2 shows the generated conceptual lattice for Table 3.

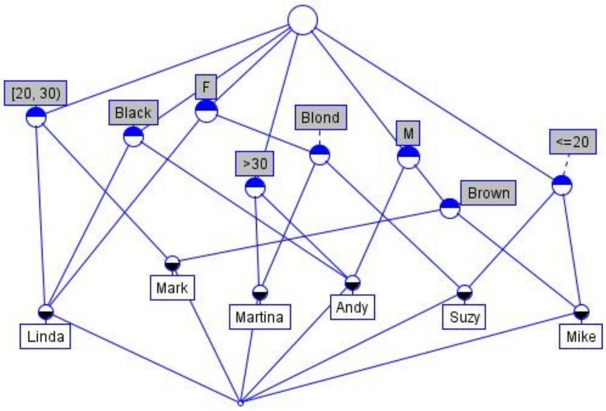


Figure 2. Conceptual lattice.

Notice that the blue nodes represent attributes, while the blue nodes represent objects. Also, the size of the blue nodes are proportional to the number of objects they contain.

Next, we present the methodology used in this project.

## 3. Methodology

The methodology used to accomplish this work includes the following steps:

- Crime data acquisition for the city of Mossoro;
- Pre-processing of the obtained data;
- Definition of criminal hot spots using Fuzzy K-means;
- Extraction of concept lattices for crime patterns using different periods of the day and types of crimes and their combinations;

- Definition of a ranking of dangerousness from the combination of the criminal hot spots and concept lattices;
- Analysis and validation of results.

Such steps are described next.

### 3.1. *Data acquisition for the city of Mossoro*

The data used in this work consists of a set of 7,486 crime occurrence records obtained from the original base that contained 43,390 records collected from December, 2011 to June, 2013. The data was provided by the Integrated Operations Centre for Public Security (Centro Integrado de Operacões em Seguranca Pública, CIOSP) of Mossoro, Brazil [b]. The data included: 4,765 records of theft, 1,171 of robbery, 229 of homicides, and 1,321 of drug traffic occurrences. The original database was structured with nine attributes described in Table 4.

| Attribute | Name | Description |
|---|---|---|
| 1 | Date | Date the occurrence was recorded |
| 2 | Hour | Hour the occurrence was recorded |
| 3 | DoW (Day of Week) | Code of the day of the occurrence |
| 4 | neighbourhood | neighbourhood where the crime occurred |
| 5 | Latitude | Latitude where the crime occurred |
| 6 | Longitude | Longitude where the crime occurred |
| 7 | Disp | Code of the occurrence dispatcher |
| 8 | Type-code | Criminal type code registered |
| 9 | Type-crime | Criminal type registered |

Table 4. Description of the attributes of occurrence records of CIOSP.

It is common sense among experts that the most important attributes to characterize a crime are its i)place, ii)date, iii)time and iv)type of crime. As previously described, with such attributes, it was possible to obtain the locations of high crime concentration, crime patterns and a ranking of districts according to the calculated level of dangerousness.

### 3.2. *Pre-processing of the obtained data*

The preprocessing of data was carried out according to the following steps:

(1) Selection of useful attributes: some attributes were discarded as they added no information, such as *dispo* and *tycod*;
(2) Removal of records with missing attribute values;
(3) Standardization of data for the ARFF format used with WEKA - Waikato Environment for Knowledge Analysis [11];

[b] `http://www.pm.rn.gov.br/Index.asp`

(4) Scaling of continuous and multivalued attributes: the scaling process transforms continuous and multivalued attributes into binary attributes for the extraction of formal concepts;
(5) Standardization of data to the format of the Conexp tool [28], which provides implementations of the algorithms to extract and display formal concepts.

Another important preprocessing task was the necessary scaling of attributes to be used with formal concepts analysis. This way, for the attributes that characterize a criminal occurrence, two values were used, 0 and 1, to represent the occurrence (or not) of the crime.

For the scaling of *Date*, two new attributes were created, one for each of the possible values of the attribute: during the week or during the weekend (see Table 5).

| **Date** | **Week** | **Weekend** |
|---|---|---|
| Week | ⊛ | |
| Weekend | | ⊛ |

Table 5. Definition attribute "'Date"' into binary values.

For the *Time* attribute, four new ones were derived: *Evening*, *Morning*, *Afternoon*, and *Night*, according to the time the crime occurred (see Table 6).

| **Time** | **Night** | **Morning** | **Afternoon** | **Evening** |
|---|---|---|---|---|
| Night (0 am to 05.59 am) | ⊛ | | | |
| Morning (6 am to 11.59 am) | | ⊛ | | |
| Afternoon (12 pm to 5.59 pm) | | | ⊛ | |
| Evening(6 pm to 11.59 pm) | | | | ⊛ |

Table 6. Definition of attribute Time in binary values.

The last attribute, type of crime (*Typ-Crime*), was scaled by creating four new binary attributes representing each of the types of crimes existing in the records: theft, robbery, homicide, and drug traffic (see Table 7).

| **TYP-CRIME** | **Theft** | **Robbery** | **Homicide** | **Drug Traffic** |
|---|---|---|---|---|
| Theft | ⊛ | | | |
| Robbery | | ⊛ | | |
| Homicide | | | ⊛ | |
| Drug Traffic | | | | ⊛ |

Table 7. Definition of attribute Typ-Crime in binary values.

Notice that in everyday language, theft and robbery are used interchangeably. However, "theft" refers to any action which involves taking any kind of property from someone without their knowledge with the intentions of depriving that person of the property permanently.

"Robbery", on the other hand, refers to taking anything valuable from another person by using threats, force, and intimidation. It may be any form of violence. The main thing about robbery is the presence of a victim.

### 3.3. *Definition of criminal hot spots using the Fuzzy K-Means algorithm*

The implementation available at KEEL [2] of the Fuzzy K-Means algorithm was used. The parameter values used were:

- Maximum number of iterations: 500;
- Number of clusters: 8 to 20.

Notice that the classic K-Means algorithm was also used to create hot spots using the Euclidean distance. However, such results are not shown as they do not present meaningful differences.

The process of defining the hot spots and the definition of the clusters are described next.

#### 3.3.1. *Definition of the clusters and experiments*

The number of clusters were defined as 8 to 20, as the police resources of the city of Mossoro in terms of vehicles vary from 8 to 20 depending on the day, time, and special events, such as public festivities and parades, which require extra vehicles and policemen. This way, all the experiments were repeated 12 times using from 8 to 20 clusters.

The second division of the experiments is related to the time of day: morning, afternoon, evening and night. The purpose of this subdivision is the possibility of obtaining different criminal hot spots for each period of the day in order to facilitate the planning of the distribution of resources, vehicles and policemen, as both vary during each period of a day. This way, all the experiments were repeated 4 times for the day, and 12 times for the number of clusters, totalling 48 different sets of clusters.

#### 3.3.2. *Extraction of latitude and longitude of the criminal records*

The addresses of the crime occurrences were used by the MarkerClusterer API of Google Maps [9] to define the latitude and longitude of each crime record. This library analyses the distance between the markers using a common formula in navigation named Haversine's formula, which measures the distance between two points from their geographical coordinates. This procedure enabled the automatic analysis of the crime patterns in the city of Mossoro by mapping and grouping them according to the types of crimes and the times they occur.

#### 3.3.3. *Extraction of concept lattices for crime patterns*

The Conexp Explorer framework [25], an open source tool implemented in JAVA was used to obtain the concept lattice models from the criminal records. The extraction of the lattices was guided by the period of the day and the neighbourhoods, using different sets of data. This way, several different lattices were extracted, as described next.

(1) A lattice with all records: a lattice with information regarding each type of crime in each time of day for each city neighbourhood (416 lattices in total);
(2) A lattice for each period of the day (morning, afternoon, evening, and night): four lattices containing information regarding the incidence of each crime in each city neighbourhood;
(3) A lattice for each type of crime: four lattices with information regarding the incidence of each type of crime for each period of time in each city neighbourhood;
(4) A lattice for each neighbourhood: 26 lattices with information regarding each type of crime in each time of day;
(5) A lattice for each period of day and for each neighbourhood: 104 lattices with information on each type of crime for each period of the day in each neighbourhood;
(6) A lattice obtained with all records for all periods of the day for each different type of crime - this model was generated using all records of the base, *i.e.*, only a criminal kind at all times of the day. These models identify the differences in rates of each crime for all periods of the day in a neighbourhood.

In total, 555 different lattices were generated and analysed. A summary of them is presented in the Results Section. It is important to outline the fact that the Conexp Explorer provides an easy to use user interface that allows the generation of several different models rapidly.

### 3.4. *Definition of a ranking of dangerousness from the combination of criminal hot spots and concept lattices*

The criminal hot spots defined with the Fuzzy K-Means algorithm were used in maps in order to allow the visualization of crime concentration in the city neighbourhoods. The data was also used n combination with a predefined formula in order to set a degree of dangerousness for each city neighbourhood.

For the definition of this formula, it was necessary to weigh each type of crime in order to allow a comparison among the different neighbourhoods. The weighting of each type of crime was defined using common sense and the opinion of experts. For instance, it is common sense that a homicide is a more severe crime than a theft. Table 8 shows the dangerousness factor we defined for each crime, which are multiplied by the number of records for each type of crime.

Using the weights above described, the crime rate for each neighbourhood is calculated according to Equation 1. Notice that the word "Weight" is abbreviated

| Crime | Factor | Weight |
|---|---|---|
| Theft (victim not present) | 1 | Theft Weight = (theft dangerousness factor) × (number of theft occurrences) |
| Robbery | 2 | Robbery Weight= (robbery dangerousness factor) × (number of robbery occurrences) |
| Homicide | 3 | Homicide Weight= (homicide dangerousness factor) × (number of homicide occurrences) |
| Drug Traffic | 4 | Drug Traffic Weight= (traffic dangerousness factor) × (number of traffic occurrences) |

Table 8. Calculation of the Danger Factor for each crime type.

to "W" for space reasons.

$$Danger\_Level = \frac{Theft\_W + Robbery\_W + Homicide\_W + Drug\_Traffic\_W}{Total\_Weight} \tag{1}$$

The actual resulting ranking is presented in Section 4.3. Next, we present and discuss the generated models.

## 4. Results

The three main contributions of this project were the i)criminal hot spots, ii) conceptual lattices, and the iii)ranking of the neighbourhoods. These models are described next.

### 4.1. *Criminal Hot Spots*

The Fuzzy K-means algorithm was used to generate clusters with a total of 104 criminal hot spot maps generated according to:

- Time of day: night, morning, afternoon and evening;
- Number of hot spots: models with 8 to 20 criminal hot spots.

Figure 3 presents a summary of the 52 generated maps of criminal hot spots for each clustering algorithm.

As previously stated, the Fuzzy K-Means algorithm was used to generate clusters with 8 to 20 centroids representing the criminal hot spots of the maps. Therefore, 13 maps were generated for each period of the day. The identification of criminal hot spots using the geographic coordinates of the crime records allowed the creation of crime maps that can be used to analyse the neighbourhoods of the city according to the types of crimes and period of the day.

For instance, Figure 4 - a shows the hot spot map obtained using 12 centroids, representing 12 criminal hot spots, for the period of the night using Fuzzy K-Means. The analysis of the map shows that the Alto de Sao Manuel neighbourhood, bottom right area of the map, presents the highest concentration of hot spots.

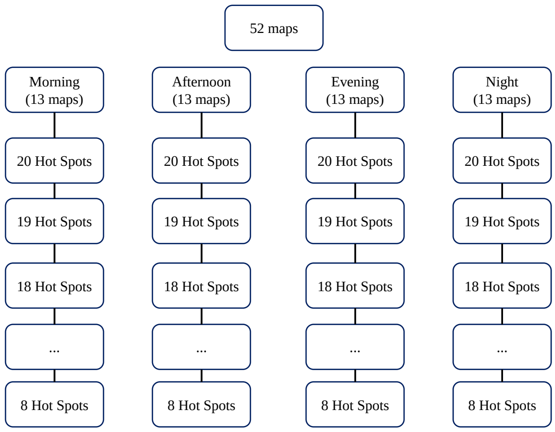


Figure 3. Criminal Hot Spots obtained with the classic and fuzzy K-Means algorithms.

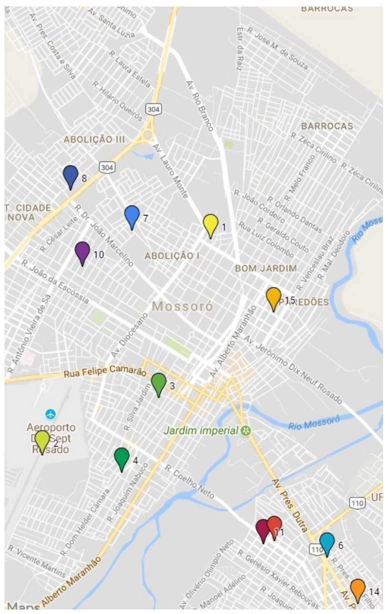


Figure 4. Criminal hot spots for the night period using Fuzzy K-Means.

Similarly, Figure 5 shows the hot spot map obtained using 12 centroids representing 12 criminal hot spots for the morning period. In this map, the centre region presents a concentration of hot spots.

Figure 6 shows the hot spot map obtained using 12 centroids, representing 12 criminal hot spots, for the period of the afternoon.

Finally, Figure 7 shows the hot spot map obtained using 12 centroids, represent-

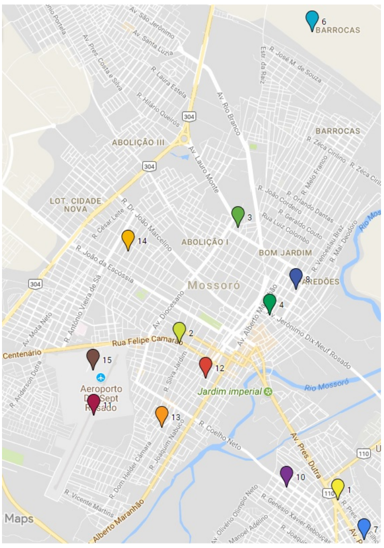


Figure 5. Criminal hot spots for the morning period using Fuzzy K-Means.

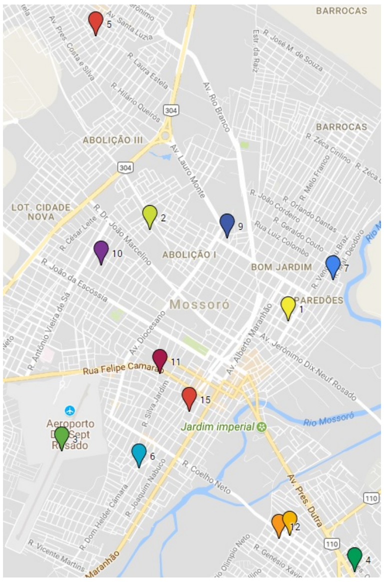


Figure 6. Criminal hot spots for the morning period using Fuzzy K-Means.

ing 12 criminal hot spots, for the period of the night.

The analysis of the four hot spot maps with 12 hot spots shows that there is a migration of crime occurrences from period to period. This information allows the definition of strategic plans for police patrol, improving the use of public resources

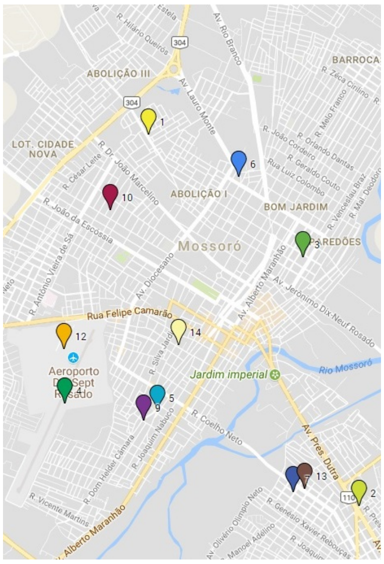


Figure 7. Criminal hot spots for the night period using Fuzzy K-Means.

and the safety in the city.

### 4.2. *Conceptual Lattice*

The use of FCA allowed the generation of a set of ten different types of models of concept lattices with information about criminal activity in Mossoro, described next:

(1) One single lattice generated using all records and all attributes;
(2) Two lattices with the distribution of crimes for each type of day (crimes during week days and during weekend days);
(3) Four lattices with the distribution of crimes for each period of the day (crimes in the morning, afternoon, evening, and night);
(4) Four lattices with the distribution of crimes for each type of crime (robbery, theft, homicide, and drug traffic);
(5) Eight lattices with the distribution of crimes for each type of crime and each type of day (crimes of robbery during week days, robbery during weekends, crimes of theft during week days, theft during weekends, crimes of homicide during week days, homicide during weekends, drug traffic crimes during week days, and drug traffic crimes during weekends);
(6) Eight lattices with the distribution of crimes for each period of the day and each type of day (crimes in the morning during the week, crimes in the morning during weekends, crimes in the afternoon during the week, crimes in the afternoon during weekends, crimes in the evening during the week, crimes in the evening during weekends, crimes at night during the week, and crimes at

night during weekends);

(7) 16 lattices with the distribution of crimes for each type of crime and each period of the day (crimes of robbery in the morning, robbery in the afternoon, robbery in the evening, and robbery at night; crimes of theft in the morning, theft in the afternoon, theft in the evening, and theft at night; crimes of homicide in the morning, homicide in the afternoon, homicide in the evening, and homicide at night; crimes of drug traffic in the morning, drug traffic in the afternoon, drug traffic in the evening, and drug traffic at night);

(8) 26 lattices with the distribution of crimes for each period of the day in each one of the 27 neighbourhoods (see Table 9);

(9) 32 lattices with the distribution of crimes for each type of crime, each period of the day, and each type of day (crimes of robbery in the morning on week days, crimes of robbery in the morning on weekends, crimes of robbery in the afternoon on week days, crimes of robbery in the afternoon on weekends, crimes of robbery in the evening on week days, crimes of robbery in the evening on weekends, crimes of robbery at night on week days, and crimes of robbery at night on weekends; crimes of theft in the morning on week days, crimes of theft in the morning on weekends, crimes of theft in the afternoon on week days, crimes of theft in the afternoon on weekends, crimes of theft in the evening on week days, crimes of theft in the evening on weekends, crimes of theft at night on week days, and crimes of theft at night on weekends; crimes of homicide in the morning on week days, crimes of homicide in the morning on weekends, crimes of homicide in the afternoon on week days, crimes of homicide in the afternoon on weekends, crimes of homicide in the evening on week days, crimes of homicide in the evening on weekends, crimes of homicide at night on week days, and crimes of homicide at night on weekends; crimes of drug traffic in the morning on week days, crimes of drug traffic in the morning on weekends, crimes of drug traffic in the afternoon on week days, crimes of drug traffic in the afternoon on weekends, crimes of drug traffic in the evening on week days, crimes of drug traffic in the evening on weekends, crimes of drug traffic at night on week days, and crimes of drug traffic at night on weekends);

(10) 26 lattices with the distribution of crimes for each neighbourhood of Mossoro (see Table 9).

Due to space limitations, next, we present and discuss some of the lattices obtained with the data from Mossoro.

Figure 7 shows the concept lattice obtained with all records: theft, robbery, homicide and traffic for all periods.

Notice that the size of the nodes is proportional to the number of involved records. Therefore, the observation of the distribution of each crime for each period of the day is straightforward. For instance, if you want to analyse the occurrence of Homicides in the Evening, you can start by following the Homicide node in the direction of the Evening node, which are connected by a larger node, whose size

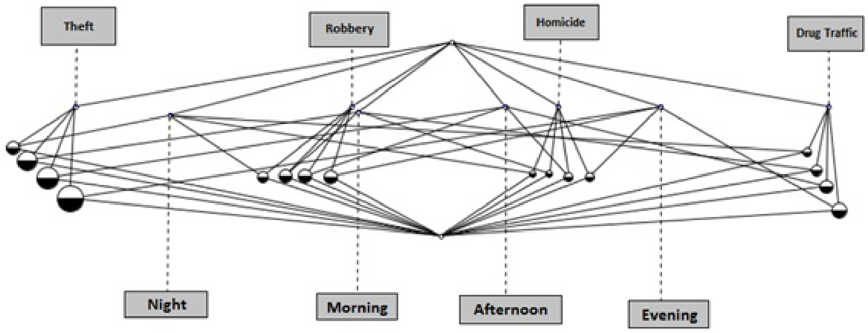


Figure 8. Conceptual Lattice showing the incidence of crimes per period of the day

is proportional to the number of occurrences of Homicide in the Evening period. Figure 7 shows that during the evening, the number of occurrences for all types of crimes is larger than for other periods of the day. Regarding Theft, the most common crime, it tends to happen more in the evening while homicides, the crime with smallest incidence, is approximately proportional in the afternoon and in the evening. Robbery, on the other hand, happens proportionally in the morning, afternoon and evening, and less at night. Drug Traffic crimes increase from the night until the evening.

Figure 9 shows the obtained conceptual lattice for all criminal records for the Alto de Sao Manoel neighbourhood during week days and weekends at night. By analysing this model, it is possible to see that, considering the proportion of days during the week (five) to the days on weekends (two), the number of occurrences is considerably high during weekends.

Figure 10 shows the concept lattice obtained for all crime types, for week days and weekends, at night, for the Abolicao neighbourhood. Notice that during the week and weekends, the most frequent crime is Theft. Another important observation is that no homicide occurrences were found in the records, which is reflected in the Lattice.

### 4.3. *Ranking neighbourhoods according to their level of dangerousness*

The identification of criminal hot spots and concept lattices allowed the definition of a dangerousness level for each neighbourhood of the city. Following this definition, the neighbourhoods were ranked from the least to the most dangerous one using a normalized value from 0 (meaning that that neighbourhood was the least dangerous) to 1 (the most dangerous neighbourhood).

In order to allow a more human interpretable and general classification of the neighbourhoods, the continuous values representing the dangerousness of each neighbourhood were translated into linguistic values. The definition of the linguistic

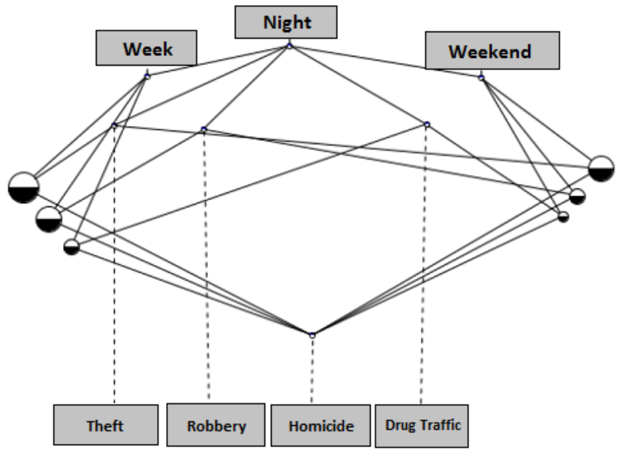


Figure 9. Conceptual Lattice showing the incidence of crimes during week days and weekends for the Alto de Sao Manoel neighbourhood at night.

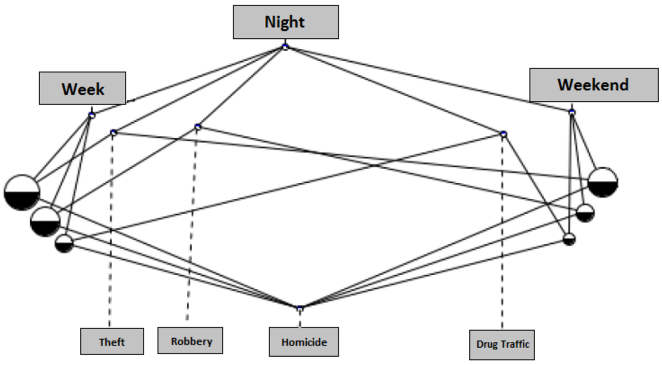


Figure 10. Conceptual Lattice showing the incidence of crime during week days and weekends at night in the Abolicao neighbourhood at night.

values was performed according to the following intervals:

- Not dangerous $[0.00, 0.15)$
- Little dangerousness: $[0.15, 0.30)$;
- Dangerous: $[0.30, 0.50)$;
- Very dangerous: $[0.50, 1.00)$.

Such intervals were defined with the support of experts in the field.

Table 9 shows the ranking of dangerousness and the corresponding linguistic values obtained for each of the neighbourhoods in Mossoro according to the proposed strategy.

| | Neighborhood | Criminality Level | Rank |
|---|---|---|---|
| 1 | Abolicao | 1.00 | Very Dangerous |
| 2 | Centro | 1.00 | Very Dangerous |
| 3 | Santo Antonio | 0.94 | Very Dangerous |
| 4 | Alto Sao Manoel | 0.82 | Very Dangerous |
| 5 | Aeroporto | 0.68 | Very Dangerous |
| 6 | Santa Delmira | 0.54 | Very Dangerous |
| 7 | Dom Jaime Camara | 0.51 | Very Dangerous |
| 8 | Nova Betania | 0.51 | Very Dangerous |
| 9 | Bom Jardim | 0.39 | Dangerous |
| 10 | Rincao | 0.39 | Dangerous |
| 11 | Belo Horizonte | 0.38 | Dangerous |
| 12 | Boa Vista | 0.37 | Dangerous |
| 13 | Planalto Treze de Maio | 0.37 | Dangerous |
| 14 | Alto da Conceicao | 0.27 | Little dangerousness |
| 15 | Doze Anos | 0.23 | Little dangerousness |
| 16 | Paredoes | 0.23 | Little dangerousness |
| 17 | Alto do Sumare | 0.20 | Little dangerousness |
| 18 | Presidente Costa e Silva | 0.20 | Little dangerousness |
| 19 | Barrocas | 0.19 | Little dangerousness |
| 20 | Pintos | 0.14 | Not dangerous |
| 21 | Redencao | 0.12 | Not dangerous |
| 22 | Ilha de Santa Luzia | 0.10 | Not dangerous |
| 23 | Bom Jesus | 0.09 | Not dangerous |
| 24 | Lagoa do Mato | 0.08 | Not dangerous |
| 25 | Dix-Sept Rosado | 0.05 | Not dangerous |
| 26 | Itapetinga | 0.01 | Not dangerous |
| 27 | Alagados | 0.00 | Not dangerous |

Table 9. Ranking of dangerousness factor.

From this set of results it is possible to define more efficient strategies against crime by optimizing the allocation of police resources, for instance. The information on the level of dangerousness is also interesting for the general public, especially for people looking to buy or rent properties in the city.

### 4.4. *Analysis and validation of results*

All results were analysed and validated by experts in the area. These public security professionals provided positive feedback and contributed with suggestions for the improvement of the models. In fact, the experts considered the positioning of the criminal hot spots to be correct, as they are set in regions they consider to be the most dangerous ones.

Regarding the lattices, the experts considered them to be quite useful for the planning of strategies to combat crime as they are simple to understand and allow quick comparisons once the size of each concept is proportional to the number of records it represents.

The experts also suggested the generation of hot spot maps and different lattices for special periods of the year when festivities occur in order to further analyse the models. For instance, during carnival, one of the most popular celebrations in Brazil, the incidence of crimes increase in general and the generated models can be useful for the planning of strategies for crime combat.

## 5. Final considerations

The analysis of large amounts of crime data is a highly relevant task for the prevention and combat of crime. This task can save public money by allowing the definition of more effective strategies for crime combat and prevention according to models induced with local customized information.

This work focuses on the use of computational intelligence techniques to allow the automatic criminal data analysis process for the city of Mossoro, Brazil. We obtained real data from police records that were used to induce models defining criminal hot spot maps and lattices with information on the incidence of crimes according to the type of crime, period of the day, neighbourhood of the city, as well as for week days and weekends.

The criminal hot spot maps were obtained using clustering analysis techniques, more specifically, the Fuzzy K-Means algorithm. Fuzzy K-means was used with the geographic coordinates of crime records to group them. The formal concept analysis theory was used to obtain conceptual lattices with information on crime incidence for a series of parameters. Such lattices provide visual models with information comparing the incidence of crimes which are highly interpretable to experts in public security. Moreover, they can be easily generated, once the data is preprocessed, for several parameters, including different types of crime, days of the week, and periods of the day.

As a last contribution, the results of the crime hot spot maps and the lattices were used to define a classification system to evaluate the level of dangerousness of each city neighbourhood of Mossoro. This classification was then used to rank the neighbourhoods of the city using linguistic values, which are more interpretable to users.

The obtained models and results were evaluated and validated by experts from the Integrated Operations Centre for Public Security of Mossoro, which provided positive feedback and have used them in their daily planing routines. They also suggested the generation of specific models for special holidays. We intend to work on such suggestions in the future. We intend to generate hot spot maps and different lattices for special periods of the year when festivities occur, such as carnival. We also intend to improve the ranking of the neighbourhoods by using fuzzy linguistic variables as inputs, in order to allow the creation of a fuzzy rule base with the support of experts to classify each neighbourhood. We also intend to collect data from other cities in order to generate similar models and have them analysed by other police department experts.

## 6. *Acknowledgments

We thank the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Brazilian National Council for Scientific and Technological Development (CNPq) for the financial support.

1. M. Agnes, editor. *Webster's New World College Dictionary*. Webster, 2015.
2. J. Alcal-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. Garca, L. Snchez, and F. Herrera. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17:255–287, 2011.
3. S. Bueno. Brazilian Institute of Geography and Statistic (in Portuguese), 2014. `http://www.ibge.gov.br/english/`.
4. M. E. Cintra. *Genetic generation of fuzzy knowledge bases:new perspectives*. PhD thesis, University of So Paulo, 2012.
5. A. Coelho and V. Furtado. Towards optimal police patrol routes with genetic algorithms. In *Lecture notes in Computer Science*, pages 485–491, 2006.
6. R. S. de Lima and S. Bueno. Brazilian forum on public safety (in Portuguese), 2014. ISSN 1983-7364.
7. A. M. G. d. Farias. Defining policing strategies to combat and prevent crimes using clustering and formal concept analysis techniques (in portuguese). Master's thesis, Post Graduation Programme in Computer Science - UERN/UFERSA, 2014.
8. W. G. Gay, T. M. Beall, and R. A. Bowers. *Integrated Criminal Apprehension Program*. Science Center, 1982.
9. GoogleApps. Markerclusterer for google maps, 2015. `https://googlemaps.github.io/js-marker-clusterer/docs/reference.html`.
10. K. Harries. Mapping crime: Principle and practice, by. Technical report, U.S. Department of Justice, 1999.
11. R. C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11(1):63–90, 1993.
12. J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation (3rd Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2006.
13. S. Institute, 2015. `http://www.mapadaviolencia.net.br/`.
14. K. Kanmehr and R. Alhajj. Effectiveness of support vector machine for crime hot-spot prediction. *Applied Artificial Intelligence*, 22:433–458, 2008.
15. Q. Kester. Criminal geographical profiling: Using FCA for visualization and analysis of crime data. *CoRR*, abs/1310.0864, 2013.
16. Q. Kester. Visualization and analysis of geographical crime patterns using formal concept analysis. *CoRR*, abs/1307.8112, 2013.
17. R. Linden. Clustering techniques. *Revista de Sistemas de Informao da FSMA (in Portuguese)*, 4:18–36, 2009.
18. M. Matteucci. A tutorial on clustering algorithms. politecnico di milano, 2013. `http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/`.
19. A. Musdholifah, S. Z. B. M. Hashim, and I. Wasito. KNN-kernel based clustering for spatio-temporal database. In *Computer and Communication Engineering (ICCCE), 2010 International Conference on*, pages 1–6. IEEE, May 2010.
20. U. D. of Justice. Mapping crime: Understanding hot spots, 2005. `http://discovery.ucl.ac.uk/11291/1/11291.pdf`.
21. D. W. Osgood. Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology*, 16(1):21–43, 2000.

22. B. T. Peixoto. *Crimina Analisys*, chapter First Module. SENASP/MJ, 2008.
23. A. Reget, editor. *Implementing Hotspot Policing: A Review of the Literature*. Portland State University, 2013.
24. S. Russel and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2 edition, 2003.
25. T. Tilley. Tool support for fca. *LNCS (LNAI)*, 2961:104–111, 2004. Springer.
26. J. N. Vidal Filho. The citizen as a voluntary human sensor in a spatially enabled society: a case study with geographic information in the area of public safety. Master's thesis, Post Graduation Programme in Computer Science - Federal University of Vicosa, 2013. (in Portuguese).
27. R. Wille. *Formal Concept Analysis. Restructuring lattice theory: an approach based on hierarchies of concepts.*, volume 5548 of *Lecture Notes in Computer Science*. Springer, 2009.
28. S. Yevtushenko. Conexp software, 2013. `http://www.sf.net/projects/conexp`.
29. P.-H. Yu and J.-G. Lay. Exploring non-stationarity of local mechanism of crime events with spatial-temporal weighted regression, 2011.
30. X. Zhang, H. Zhiang, R. Li, and Z. Zheng. Detecting and mapping crime hot spots based on improved attribute oriented induce clustering. In *IEEE International Conference on Geoinformatics*, pages 1–5, 2010.