

Spatial association discovery process using frequent subgraph mining

Giovanni Daián Rottoli¹, Hernán Merlino²

¹Universidad Nacional de La Plata, Argentina

¹Universidad Tecnológica Nacional, Argentina

^{1,2}Information Systems Research Group, National University of Lanús, Buenos Aires

Article Info

Article history:

Received Aug 10, 2019

Revised Mar 10, 2020

Accepted Apr 3, 2020

Keywords:

Frequent subgraph mining

SARM

Spatial association mining

Spatial data mining

Spatial knowledge discovery

ABSTRACT

Spatial associations are one of the most relevant kinds of patterns used by business intelligence regarding spatial data. Due to the characteristics of this particular type of information, different approaches have been proposed for spatial association mining. This wide variety of methods has entailed the need for a process to integrate the activities for association discovery, one that is easy to implement and flexible enough to be adapted to any particular situation, particularly for small and medium-size projects to guide the useful pattern discovery process. Thus, this work proposes an adaptable knowledge discovery process that uses graph theory to model different spatial relationships from multiple scenarios, and frequent subgraph mining to discover spatial associations. A proof of concept is presented using real data.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Giovanni Daián Rottoli,

Departamento de Ingeniería en Sistemas de Información,

Universidad Tecnológica Nacional, F.R. Concepción del Uruguay,

676 Ing. Pereira Street, Concepción del Uruguay (3260), Entre Ríos, Argentina.

Email: rottolig@frcu.utn.edu.ar

1. INTRODUCTION

Spatial knowledge discovery aims to find useful and novel patterns in spatial datasets to support decision-making in a particular problem domain [1]. Among all the possible patterns to discover, spatial associations are one of the most commonly used today in multiple fields such as climatology, geography, geology, criminology and ecology, among many others. They are comprised of predicates that involve spatial objects along with spatial and non-spatial relationships between those objects [2]. There are many challenges associated with the characteristics of spatial data that make this data mining task more complicated, such as the spatial dependency data attributes, the multiplicity of spatial data representation models, the spatial relations between data objects and some particular spatial properties such as spatial autocorrelation and spatial heterogeneity [3].

Multiple algorithms have been developed for association pattern mining that can be used. Each of these algorithms, in general, aims to solve particular concerns about the aforementioned challenges. The selection of a proper algorithm has become an arduous activity due to the growing number of new alternatives and their variants, specially to inexperienced users. Thus, it is necessary to provide a new process for small or medium-size application domains, one that is easy to implement and flexible enough to be adapted to multiple contexts. Consequently, this paper proposes a new process for association mining discovery from spatial data

that utilizes graph theory to model spatial objects and the relations between them and frequent subgraph mining to find the substructures with a high repetition rate inside the general graph. These substructures correspond to association patterns. The proposal is a new alternative to model complex situations from a particular problem domain, but not replace or improve results from the algorithms in the state-of-the-art, however it provides a road map to initially address a problem. The rest of this paper is arranged as follows: section 2. on the characteristics of spatial data; section 3. contains association patterns and their characteristics regarding spatial data; section 4. includes the proposed process for discovery of spatial associations; a proof of concept using real world data is shown in section 5. Lastly, section 6. contains conclusions and future works .

2. SPATIAL DATA

Spatial data is a particular type of dependent data. Formally, a spatial database D is a set of spatial records $D = \{\vec{T}_1, \vec{T}_2, \dots, \vec{T}_d\}$ with $T_i = \{S_i^1, S_i^2, \dots, S_i^m, X_i^1, X_i^2, \dots, X_i^n\}$, where each S_i^k is a spatial attribute that stores values about the spatial contexts, and each X_i^l is a non-spatial attributes with values measured at particular locations [3, 4]. The non-spatial attributes may be numerical or categorical according to the problem domain and the spatial attributes may be specified as coordinates or places (e.g. city name or state code). Additionally, there are three basic types of spatial objects: points, used to model specific punctual locations in the space; lines, used to model linear extensions such as rivers or roads; and polygons, used to represent objects that have a two-dimensional extension in the space, such as regions or states.

The dependence of non-spatial attributes on spatial ones means that different implicit spatial relations can be extracted from data. Let D be a spatial database, a relation $R \subseteq D^2$ is called spatial if and only if it is defined through a binary predicate $P(x, y) | x, y \in D$ that involves the spatial attributes from the spatial objects x and y . For example, the spatial relation $N \subseteq D^2$, with $x, y \in D$, defined by the predicate shown in (2), is the neighborhood relations between two spatial points using euclidean distance: $xNy \iff Dist(x, y) < \lambda$, $\lambda \in R^+$

These relations can be classified as *geometric*, if they are related to the principles of euclidean geometry (e.g. neighbouring relationships); *directional*, when they refer to relative spatial orientations (e.g. above, below, north, east); *topological*, if they are independent from the concepts of distance and direction and are not affected by spatial transformations such as rotation or translation (e.g. intersect, inside), or *hybrid*, if they are related to two or more of the aforementioned types of properties. These relationships can be calculated using different methods depending on the problem domain and the class of spatial data used: points, lines or polygons [5, 6].

On the other hand, two properties are derived from spatial dependence: spatial autocorrelation, i.e., observations of spatially distributed random variables are not location-independent, and spatial heterogeneity, i.e., patterns found in some region of the space may not have the same support in other region. Spatial autocorrelation refers to the particularity of spatial data to not be distributed independently throughout the space. The distribution depends on the characteristics of the data points, the characteristics of the underlying space or the spatial neighboring relationships. For example, churches tends to be located near public squares or animal tends to travel to locations that contain their food sources [7]. Spatial heterogeneity is related to spatial autocorrelation. This phenomenon describes the local nature of spatial patterns, which are subordinated to some specific locations. Thus, a spatial pattern, such as association rules, may have a high support value in a region and a low support value in a different one. This phenomenon is also known as Simpson's paradox [8]. All these particular characteristics make knowledge extraction from spatial data become a complex activity which not only has to consider patterns between data records, but also the implicit relationships between spatial objects.

3. SPATIAL ASSOCIATIONS

One of the most common patterns to find in data is the association pattern. An association pattern P is defined as an n-ary predicate $P = (p_1, p_2, \dots, p_n)$ with a high probability of occurrence in the dataset. Its classic application is the supermarket basket analysis to discover whether or not there is some correlation between items that are bought together. An association pattern is referred to as spatial if at least one of its atomic predicates p_k involves a spatial relationship between its variables [2]. For example, in a city C , churches and public squares tend to be neighbors: $City(C) \wedge Church(X) \wedge PublicSquare(Y) \wedge Inside(X, C) \wedge Inside(Y, C) \wedge Neighbors(X, Y)$

As shown in the previous example, $Inside(X, C)$, $Inside(Y, C)$ and $Neighbors(X, Y)$ are spatial

predicates related to topological and geometric relationships. Many different relations must be taken into consideration at the same time to find useful spatial associations. Also, these relations must be calculated in local contexts, due to the aforementioned Simpson's Paradox.

Multiple efforts have been made in order to find spatial association patterns in spatial databases: [7] proposes a method for spatial association mining that consider spatial autocorrelation by using a cell structure; [9] focuses on the problem of rule extraction from spatial data with crisp condition attributes and fuzzy decisions. A rough-fuzzy set based rule extraction model is used to deal with both fuzziness and roughness; [10] combines and extend techniques developed in both spatial and fuzzy data mining to deal with the uncertainty found in typical spatial data. This proposal uses fuzzy logic to get relevant information from transition areas between spatial neighborhoods to spatial association mining and for spatial relationships modelling; [11, 12] propose an algorithm for local patterns discovery considering spatial heterogeneity that incorporates a novel spatial metric for support evaluation based on event density in a particular area; [13] presents a specially designed algorithm to discover spatial associations related to *El Niño* Southern Oscillation (ENSO); [14] applies an algorithm that explores multiple spatial objects hierarchies; [15] uses A-Priori-based approaches to find spatial association rules; [6, 16] propose using Inductive Logic Programming (ILP) for reach this data mining purpose by modelling and stracting high support spatial relations from spatial data. [17] worked with meta-heuristics such as genetic algorithms and evolutionary programming; [18] suggested a data-transformation approach before using traditional association rule mining algorithms; [19] introduced non-trivial structures such as graphs for spatial relationship representation; among others.

Because of this variety of spatial data mining approaches for association discovery, it is difficult to select a proper algorithm or method to be used in small knowledge discovery application contexts. Because of this, a unified and general process is required to deal with the aforementioned problems and it has to be flexible enough to be adapted to multiple particular situations and easy to implement.

4. SPATIAL ASSOCIATION DISCOVERY PROCESS

This work describes a new process for spatial association extraction considering the possibility of having multiple relationships between spatial objects of any kind (i.e. points, lines, polygons), and considering the spatial autocorrelation and spatial heterogeneity. This process is designed as a first approach to get spatial association knowledge from data in particular contexts easy to implement in small or medium-size projects. The process Figure 1 is divided into 5 main steps: data preparation (section 4.1.), neighborhood definition (section 4.2.), modelling of spatial relationships using graphs (section 4.3.), frequent subgraph mining (section 4.4.) and evaluation of results (section 4.5.).

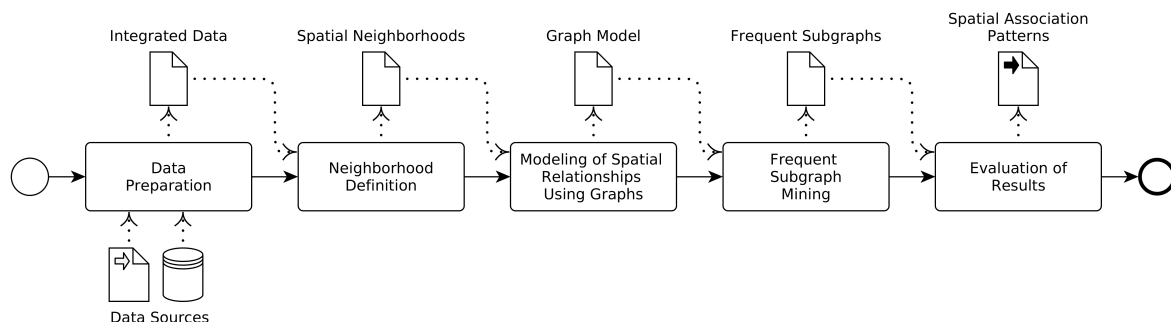


Figure 1. Spatial association discovery process

4.1. Data preparation

The proposed process starts with a spatial data preparation step. It is necessary to codify the various spatial datasets obtained from different sources in different formats, in order to enable the extraction of relations between all the data instances in later steps. In general terms, it is not uncommon to have multiple spatial objects layers, each of them with a particular representation type and related to a particular scenario from the problem domain. On the other hand, two types of datasets must be considered: target datasets, with objects directly

related to the problem domain that are going to be present in every association pattern, and relevant datasets, that may or may not be related to the target datasets, but add important information that may be useful for decision making [20].

These data must be prepared by cleaning errors, solving inconsistent and null values, and dealing with outliers. New attributes or even new data objects could be generated using the input data. This step requires considerable effort and may require many iterations. Thus, it is advisable to implement the process using a proper methodology such as CRISP-DM [21].

4.2. Neighborhood definition

As mentioned before, a particular spatial association pattern may have a higher occurrence probability in some regions and lower probability in others [8]. For this reason it is preferred to search for this kind of pattern locally. For this, we propose defining partitions of the dataset, called *neighborhoods* in this context, and the subsequent execution of the association pattern search algorithm on each of them.

These neighborhoods can be defined beforehand using knowledge from the problem domain, or using spatial clustering techniques. Using density-based or distance-based spatial clustering algorithms [22–24] is suggested due to the First Law of Geography, which states that spatial objects located together are more closely related than those that are far away from each other [25, 26]. Nonetheless, there is an issue to consider in this step: the limits between neighborhoods may add important information for spatial association mining. Thus, the use of fuzzy clustering techniques or flexible boundaries models may be desirable.

4.3. Modelling of spatial relationships using graphs

Now, we have to calculate the spatial relations between the target data instances and the instances of the relevant dataset from each neighborhood. Depending on the problem domain, different types of spatial relations can be calculated: euclidian, topological, directional or hybrid relationships, as mentioned above [6]. This might be a step with a high computational cost.

Graph theory is proposed to model the spatial relationships due to its close relation with first order logic and the pattern to find [16]. Graphs are discrete structures consisting of vertices and edges that connect these vertices. There are different kinds of graphs, depending on whether edges have directions (digraphs), whether multiple edges can connect the same pair of vertices (multigraphs), and whether loops are allowed.

Formally, a *simple graph* $G = (V, E)$ consists of V , a nonempty set of vertices (or nodes) and E , a set of edges. Each edge has two vertices associated with it, called its endpoints. An edge is said to connect its endpoints. To relate each edge to its endpoints, a function $\phi : E \rightarrow \{v1 \in V, v2 \in V\}$, called incidence function, is used. A multigraph, on the other hand, is a graph where multiple edges can exist associated with the same endpoints. Additionally, each vertex and each edge can be labeled with data related to the represented object. This structure can be adapted to multiple scenarios and multiple efficient algorithms can be used to extract valuable information such as maximum cliques [27].

In the context of this work, multigraphs are used to model spatial objects as vertices and the relations between them as edges. A small example can be seen in Figure 2 (a). Two sets of labels and two extra functions to assign those labels to the vertices and edges are needed. So, let G be a multigraph without loops $G = (V, E, L, K, \phi, \iota, \kappa)$ where: V is the vertex set of G , which corresponds to the spatial objects from the datasets; E is the edge set that corresponds to each calculated relationship between the spatial objects; L is the vertex label set with the characteristics of the spatial data objects; K is the edge label set, with the characteristics of each spatial relation; $\phi : E \rightarrow \{x \in \mathcal{P}(V) / |x| \leq 2\}$ is the incidence function; $\iota \subseteq V \times L$ and $\kappa \subseteq E \times K$ are labeling relations.

The aforementioned structure makes it possible to model multiple different relationships with the same endpoints labeled with different attributes. Also, many attributes of spatial objects could be taken into consideration. Additionally, it must be noted that loops (i.e. edges with only one endpoint) are not considered because their lack of semantics in this context (there are not spatial relationships that involves only one spatial object). Fuzzy logic could also be a valuable tool to model the spatial relationships, if the situation requires it [10]. More information about fuzzy logic this can be found in [28]

4.4. Frequent subgraph mining

To extract spatial associations with a high probability of occurrence, frequent subgraph mining is proposed to be used for each modeled graph. Given a multigraph $G = (V, E, L, K, \phi, \iota, \kappa)$ like the one described in the previous section, the frequent subgraph mining problem in a single multigraph is finding recurring subgraph

$G_i \subset G$, or in other words, a subgraph that has multiple instances in the original graph Figure 2 (b). It must be noted that two graphs are isomorphic if all of their vertices and edges are shared including its labels. These frequent subgraphs represent the relationships between spatial object types that take place in the space with a high occurrence probability.

Multiple algorithms have been designed for frequent subgraph mining in a single big graph, calculating the relevance of a pattern in different ways. Some well-known examples of this are IncGM+, FSSG, SUBDUE, among others [29, 30]. A set of frequent subgraphs for each neighborhood is obtained as a result of this step and must be analyzed to obtain useful knowledge for decision-making.

4.5. Evaluation of results

In the final step, frequent subgraphs translated into n-ary predicates that represent trivial information (non-novel patterns) must be filtered. The support and confidence measures can be extracted, selecting the metrics that the decision-maker consider to be more appropriate. This activity could be performed automatically or manually by an analyst with knowledge about the problem domain with help from an expert.

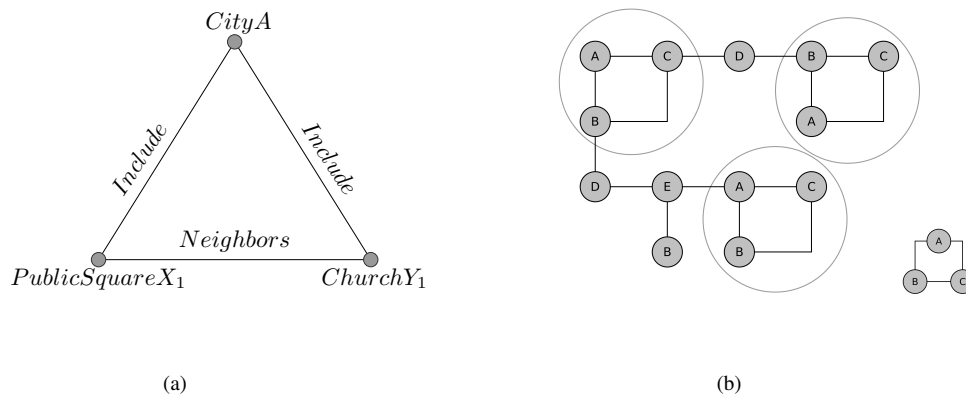


Figure 2. (a) Simplified example of spatial relationship modelling using a simple graph. (b) Example of frequent subgraph (bottom) found in a simple graph without labels in the edges.

5. PROOF OF CONCEPT

The proof of concept presented in this section is intended to show how the proposed process works, implemented by different programming and data mining tools. The data used in this example consists of 10 data files containing the location of facilities in Buenos Aires (Argentina) and its surroundings. These facilities include libraries(74), clinics(63), post offices(55), sports halls(50), nightclubs(41), schools(107), gas stations (97), churches (125), museums(37) or police stations (93).

For each of them, in the preparation step of the proposed process, the data files were integrated into a single data file of spatial points using QGIS (<http://qgis.org/>). Each spatial point is comprised of two spatial attributes, Latitude and Longitude, and one non-spatial attribute, the type of building from the previous list. After that, only the points that are located outside Buenos Aires limits were filtered to reduce the search space, leaving 742 spatial points Figure 3 (a), (orange). Then, in the neighborhood definition step, the HDDBSCAN clustering algorithm [31] from the 'dbscan' library from R programming language was used on the spatial data attributes to generate two neighborhoods with a minimum number of points equal to 50 in each of them Figure 3 (a), (blue). Only two neighborhoods were used because of explanatory purposes.

In the next step, for each of the generated neighborhoods, a geometric relationship between their data points was extracted forming a graph with vertices labeled with the type of facility related to each data point and edges labeled with the sentence "close to" if the adjacent points were less than 150 meters away from each other (this value was selected for illustrative purposes only). Thus, two graphs were created: one with 71 vertices and 45 edges in neighborhood 1, and another with 15 vertices and 11 edges in neighborhood 2.

To obtain the frequent subgraphs of each of the generated graphs, SUBDUE algorithm was used via its implementation in Subdue Graph Miner Software, using the compression rate as support measure.

The result was a subgraph as shown in Figure 3 (b), with a compression rate of 15.5% in neighborhood 1, which was translated into the predicate $\text{Post office}(x_1) \wedge \text{Nightclub}(x_2) \wedge \text{Close to}(x_1, x_2)$ and two subgraphs in neighborhood 2, both with a compression rate of 27.2% that was translated into the predicates $\text{Clinic}(x_1) \wedge \text{Post office}(x_2) \wedge \text{Close to}(x_1, x_2)$ and $\text{Post office}(x_1) \wedge \text{Sport hall}(x_2) \wedge \text{Close to}(x_1, x_2)$

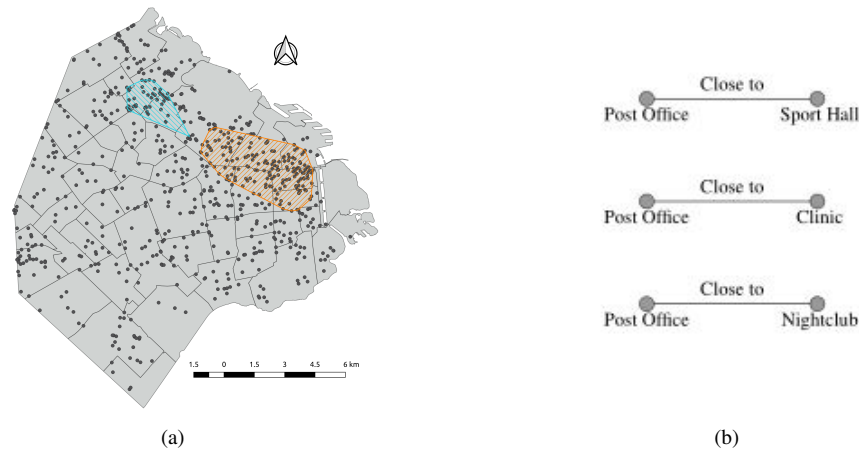


Figure 3. (a) Spatial neighborhoods generated for the proof of concept using HDBSCAN algorithm; (b) Results of the proof of concept.

5.1. Discussion

The contributions of the proposed process are, firstly, the possibility of adapting it to multiple scenarios, due to its flexible underlying structure being based on graphs. Some of the aforementioned methods use flexible structures too [6, 16] but the complexity of these methods increases because of the use of techniques based on Logic Programming. On the other hand, some other methods do not take into account complex patterns [19]. Furthermore, the possibility of including valuable information related to the data objects and the spatial relations by using labels in the graph representation is also considered. Generally, the data structures involved do not take into account complex data associated to the spatial relations between spatial data.

In relation to the above, the proposed process considers spatial phenomena such as autocorrelation and heterogeneity, by using spatial neighborhoods. Some alternatives such as [7] considering spatial autocorrelation but not considering spatial heterogeneity or complex data relationships. In most of the cases studied, these characteristics are present due to their relevance in data mining.

Also, related to this, the proposed process allows its implementation by using existing tools such as frequent subgraph mining algorithms and clustering algorithms. Some of the state-of-the-art alternatives include very flexible and powerful strategies, but implementation is hard, making them not suitable for application in small or medium size projects [6, 9, 16, 19]. Lastly, the high adaptability of the procedure is a desired characteristic due to the possibility of selecting among many algorithms for the implementation of each step. Usually, the state-of-the-art methods propose a single alternative for its execution.

6. CONCLUSION

This work describes a knowledge discovery process called for extraction of spatial associations. The process is flexible enough to take into account multiple and varied spatial relationships between spatial objects of any kind, using a graph structure to model them. Heterogeneity and autocorrelation phenomena are also considered, defining neighborhoods where the search process is performed to find this class of regularity. The solution was designed to initially approach to this data mining task without worrying too much about particular characteristics of data mining algorithms. In a large-scale project, this process could guide the selection of specific methods based on the results obtained in first iterations of an incremental methodology. A proof of concept is presented as well, using real data to illustrate how the process is implemented using different programming and data mining tools in each of the proposed steps.

In future works, the research will be focused on implementation strategies according to the problem domain for each of the steps of the process, in order to decrease computational execution time when dealing with large amounts of spatial objects and spatial relationships. Also, fuzzy methods will be considered for relation modelling and neighborhood definition.

ACKNOWLEDGEMENT

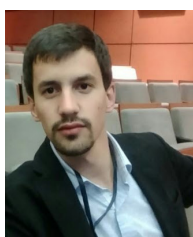
The research presented in this paper was partially funded by the PhD Scholarship Program to reinforce R&D&I areas (2016-2020) of the Universidad Tecnológica Nacional and the Research Project 80020160400001 LA of National University of Lanús. The authors also want to extend their gratitude to Kevin-Mark Bozell Poudereux, for proofreading the translation.

REFERENCES

- [1] R. Garcia-Martinez, P. Britos, and D. Rodriguez, "Information mining processes based on intelligent systems," *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 402-410, 2013.
- [2] K. Koperski and J. Han, "Discovery of spatial association rules in geographic information databases," *International Symposium on Spatial Databases*, pp. 47-66, 1995.
- [3] Y. Leung *et al.*, "Knowledge discovery in spatial data," *Springer*, 2010.
- [4] C. C. Aggarwal, "Data mining: The textbook," *Springer*, 2015.
- [5] R. Agrawal, *et al.*, "Fast algorithms for mining association rules," *Proc. 20th int. conf. very large data bases*, vol. 1215, pp. 487-499, 1994.
- [6] A. Appice, M. Ceci, A. Lanza, F. A. Lisi, and D. Malerba, "Discovery of spatial association rules in geo-referenced census data: A relational mining approach," *Intelligent Data Analysis*, vol. 7, no. 6, pp. 541-566, 2003.
- [7] J. Chen, "An algorithm about association rule mining based on spatial autocorrelation," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, no. B6b, pp. 99-106, 2008.
- [8] E. H. Simpson, "The interpretation of interaction in contingency tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 13, no. 2, pp. 238-241, 1951.
- [9] H. Bai, Y. Ge, J. Wang, D. Li, Y. Liao, and X. Zheng, "A method for extracting rules from spatial data based on rough fuzzy sets," *Knowledge-Based Systems*, vol. 57, pp. 28-40, 2014.
- [10] R. Ladner, F. E. Petry, and M. A. Cobb, "Fuzzy set approaches to spatial data mining of association rules," *Transactions in GIS*, vol. 7, no. 1, pp. 123-138, 2003.
- [11] Z. Sha and X. Li, "Mining local association patterns from spatial dataset," *Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 3, pp. 1455-1460, 2010.
- [12] Z. Sha, X. Tan, and Y. Bai, "Localized spatial association: A case study for understanding vegetation successions in a typical grassland ecosystem," *Geo-Informatics in Resource Management and Sustainable Ecosystem*, pp. 33-45, 2015.
- [13] X. Cunjin and L. Xiaohan, "Novel algorithm for mining ENSO-oriented marine spatial association patterns from raster-formatted datasets," *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, pp. 1-15, 2017.
- [14] A. Salleband and C. Vrain, "An application of association rules discovery to geographic information systems," *European Conference on Principles of Data*, pp. 613-618, 2000.
- [15] S. S. U. Sutjipto, I. S. Sitanggang, and B. Barus, "Potential usage estimation of ground water using spatial association rule mining," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 15, no. 1, pp. 504-511, 2017.
- [16] D. Malerba, F. Esposito, F. A. Lisi, and A. Appice, "Mining spatial association rules in census data," *Research in Official Statistics*, vol. 5 no. 1, pp. 19-44, 2003.
- [17] A. H. Goudarzi and N. Ghadiri, "A hybrid spatial data mining approach based on fuzzy topological relations and moises evolutionary algorithm," *Artificial Intelligence, Cornell University* 2017.
- [18] I. Lee, "Mining multivariate associations within gis environments," *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 1062-1071, 2004.

- [19] H. Yang, S. Parthasarathy, and S. Mehta, "Mining spatial object associations for scientific data," *II International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 902-907, 2005.
- [20] V. Bogorny, P. M. Engel, and L. O. Alvares, "Geoarm: an interoperable framework to improve geographic data preprocessing and spatial association rule mining." *SEKE*, pp. 79-84, 2006.
- [21] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pp. 29-39, 2000.
- [22] J. Sander, *et al.*, "Density-based clustering in spatial databases: The algorithm gdbscan and its applications," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 169-194, 1998.
- [23] Y. Zhu, K. M. Ting, and M. J. Carman, "Density-ratio based clustering for discovering clusters with varying densities," *Pattern Recognition*, vol. 60, pp. 983-997, 2016.
- [24] A. Sharma, R. Gupta, and A. Tiwari, "Improved density based spatial clustering of applications of noise clustering algorithm for knowledge discovery in spatial data," *Mathematical Problems in Engineering*, vol. 2016, 2016.
- [25] W. R. Tobler, "Cellular geography," *Philosophy in geography*, pp. 379-386, 1979.
- [26] J. Duan, L. Wang, and X. Hu, "The effect of spatial autocorrelation on spatial co-location pattern mining," *International Conference on Computer, Information and Telecommunication Systems*, pp. 210-214, 2017.
- [27] G. D. Rottoli, H. Merlino, and R. Garcia-Martinez, "Co-location rules discovery process focused on reference spatial features using decision tree learning," *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 221-226, 2017.
- [28] D. J. Dubois, "Fuzzy sets and systems: theory and applications," *Academic press*, vol. 144, 1980.
- [29] E. Abdelhamid, M. Caim, M. Sadoghi, B. Bhattacharjee, Y. Chang, and P. Kalnis, "Incremental frequent subgraph mining on large evolving graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2710-2723, 2017.
- [30] D. Kavitha, V. Kamakshi, and J. Murthy, "Finding frequent subgraphs in a single graph based on symmetry," *International Journal of Computer Applications*, vol. 146, no. 11, pp. 0975-8887 2016.
- [31] R. J. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical density estimates for data clustering, visualization, and outlier detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 1, pp. 5, 2015.

BIOGRAPHIES OF AUTHORS



Giovanni Daián Rottoli is a researcher at the Computational Intelligence and Software Engineering Research Group (GIICIS) from the National University of Technology (Argentina). He holds a Bachelor's Degree in Information Systems from the aforementioned university (2015). He is currently a Ph.D. candidate in Computer Science at the National University of La Plata (Argentina). He works as an associate professor of Discrete Mathematics and Data Science at the National University of Technology (Argentina). His research is focused in the fields of spatial data mining and knowledge discovery, artificial intelligence and search-based software engineering. Further info can be found on his profile: <http://www.frcu.utm.edu.ar/giicis/rottolig/>



Hernán Merlino is the head of the Advanced Information Systems Laboratory at Buenos Aires University (Argentina) and the head of the Artificial Intelligence Laboratory at the National University of Lanús (Argentina). He is a fellow of the Gas and Petroleum Institute at Buenos Aires University. He holds a Bachelor's Degree in Information Systems from the University of Belgrano (Argentina), a Master's Degree in Software Engineering from the Computer Science Department of the Polytechnic University of Madrid (Spain), and a Ph.D. in Information Sciences from the National University of La Plata (Argentina). He works as a tenured full professor in graduate and postgraduate courses at Buenos Aires University, Austral University and National University of Lanús. His research interests are: artificial intelligence, data mining, and blockchain technologies. In the professional field, he works as a Scientific Research Director in an Artificial Intelligence, Data Science, Blockchain & Smart Contracts company in Argentina.