

Proceso para el Descubrimiento de Asociaciones Espaciales Mediante Subgrafos Frecuentes

Giovanni Daián Rottoli^{1,2,3}[0000-0002-7623-2591], Hernán Merlino³, and Ramón García-Martínez^{3†}

¹ Programa de Doctorado en Ciencias Informáticas, Facultad de Informática, Universidad Nacional de La Plata (UNLP). La Plata, Buenos Aires, Argentina.

² Programa de Beca Doctoral para el Fortalecimiento de Áreas I+D+i. Universidad Tecnológica Nacional (UTN). Argentina
rottolig@frcu.utn.edu.ar

³ Grupo de Investigación en Sistemas de Información (GISI), Universidad Nacional de Lanús (UNLa). Lanús, Buenos Aires, Argentina.
hmerlino@unla.edu.ar

Resumen Las asociaciones espaciales constituyen uno de los tipos de regularidades más utilizadas para la inteligencia de negocio en relación a los datos espacialmente referenciados. Debido a las particularidades de este tipo de información, distintos acercamientos han sido propuestos a lo largo del tiempo para el minado de las mismas, sin embargo, no existe una forma definida de realizar esta actividad que sea adaptable a las situaciones particulares de los distintos problemas de negocio que se presentan. Por este motivo se propone en el presente trabajo un proceso de explotación de información que permita modelar distintos escenarios y relaciones espaciales haciendo uso de la teoría de grafos, y la búsqueda de patrones de asociación mediante el uso de algoritmos de búsqueda de subgrafos frecuentes. Una prueba de concepto de la propuesta utilizando datos reales es llevada a cabo para ilustrar la implementación del mismo.

Keywords: Asociación Espacial, Minería de Datos Espaciales, Explotación de Información Espacial, Subgrafos Frecuentes

1. Introducción

La explotación de información espacial busca encontrar patrones interesantes y novedosos en un conjunto de datos espacialmente referenciados para así apoyar la toma de decisiones estratégicas de una organización [1]. Entre estos patrones interesantes, las asociaciones espaciales constituyen uno de los más utilizados [2], sin embargo, debido a las características de este tipo particular de información, diferentes desafíos se presentan [3].

En primer lugar, $D = \{\{s_j\}, \{v_k\}\}$ es un dato espacial si está compuesto por dos clases de atributos, un conjunto de n atributos espaciales $\{s_j\}, j = 1..n$, y un conjunto de m atributos no espaciales $\{v_j\}, k = 1..m$. Los últimos mencionados almacenan mediciones sobre el dominio del problema, mientras que los primeros vinculan estas mediciones con un lugar específico del espacio [4].

A partir de esta dependencia espacial de los atributos se pueden extraer relaciones entre las distintas instancias de los datos tales como relaciones geométricas basadas en la distancia euclídeana, como la vecindad o la lejanía, relaciones topológicas, como las intersecciones e inclusiones, relaciones de ubicación, como que un objeto se encuentre al norte o debajo de otro, y relaciones mixtas como la perpendicularidad o el paralelismo.

Estas relaciones se calculan de diversas maneras en función del dominio del problema y del tipo de dato espacial que estemos utilizando: puntos, líneas o regiones. En consecuencia, las asociaciones en el espacio no solo contemplarán los valores de los atributos no espaciales, sino que además deberán considerar las relaciones entre los objetos espaciales, a diferencia de los patrones de asociación transaccionales tradicionales [5, 6].

Por otro lado, es necesario considerar además el fenómeno de la heterogeneidad espacial: un patrón espacial puede tener distinta probabilidad de ocurrencia en distintas ubicaciones del espacio, por lo que tiene sentido la búsqueda de patrones locales [7]. De esta forma, las asociaciones deberán ser analizadas particularmente en secciones del espacio estudiado.

Todas estas características deben ser tenidas en cuenta para encontrar patrones de asociación en datos espacialmente referenciados, sin embargo, los distintos trabajos realizados hasta la fecha solucionan los distintos problemas que se presentan de manera aislada: Saleb y Christel (2000) [8] proponen la consideración de múltiples jerarquías de datos espaciales; Sutjipto y otros [9] utilizan acercamientos basados en el algoritmo A-Priori, Malerba y Lisi (2001) [10], Malerba y Esposito (2002) [11] y Appice y otros (2003) [6] proponen la utilización de algoritmos basados en Programación Lógica Inductiva, mientras que Ladner y otros (2003) [12], y Bai y otros (2014) [13] incorporan el uso de lógica difusa (*fuzzy logic*) para el modelado de las relaciones entre los objetos espaciales, y Zhang (2017) [14] y Goudarzi y Ghadiri (2017) [15] la utilización de metaheurísticas como algoritmos genéticos y programación evolutiva. Por otro lado, Lee (2004) [16] propone un acercamiento basados en transformación de datos y Yang y otros (2005) [17] incorpora el uso de estructuras como grafos. Por último, distintos fenómenos son considerados: la localidad de los patrones por Ding y otros (2006, 2011) [18, 19] y Sha y otros (2010, 2015) [20, 21], y la autocorrelación espacial, el efecto vecindario y el efecto derrame por Dao y Thill [22].

Esta variedad de soluciones hace que aumente la complejidad de modelar el problema correctamente e implementar de forma adecuada una solución para el mismo. Por este motivo, es necesario un proceso para la búsqueda de patrones de asociación que permita el modelado de distintas relaciones espaciales simultáneamente entre distintas clases de objetos espaciales, la posibilidad de buscar los patrones localmente, y la búsqueda de patrones alrededor de objetos espaciales de referencia, para enfocar la búsqueda según el problema considerado.

Para esto, se presenta el proceso de explotación de información propuesto analizando cada una de sus etapas (sección 2) y se realiza una prueba de concepto del mismo sobre datos reales para ilustrar la implementación del mismo (sección 3). Finalmente se presentan conclusiones (sección 4).

2. Proceso para el Descubrimiento de Asociaciones Espaciales

Un patrón de asociación consiste en un predicado compuesto que posee una alta probabilidad de ocurrencia. Una asociación es llamada espacial si por lo menos uno de sus predicados simples componentes involucra una relación espacial entre objetos espaciales. Por ejemplo, sean A una ciudad perteneciente al conjunto de datos, X una iglesia ubicada en A , Y una plaza ubicada en A , entonces se verifica que en A , las iglesias se encuentran cerca de las plazas:

$$EsCiudad(A) \wedge EsIglesia(X) \wedge EsPlaza(Y) \wedge Incluido(X, A) \wedge Incluido(Y, A) \wedge Vecinos(X, Y)$$

Como se puede observar en este predicado, $Incluido(X, C)$, $Incluido(Y, C)$ y $Vecinos(X, Y)$ son predicados espaciales relacionados con relaciones topológicas y geométricas. Múltiples relaciones distintas deben ser tenidas en cuenta para encontrar asociaciones complejas que resulten interesantes para la inteligencia de negocio. Además, estas relaciones deben ser calculadas localmente debido al fenómeno ya mencionado de la heterogeneidad espacial [7].

En el presente trabajo, en consecuencia, se introduce un proceso de explotación de información que permite extraer asociaciones espaciales contemplando la posibilidad de incorporar múltiples relaciones espaciales, la heterogeneidad espacial, y el poder limitar la búsqueda alrededor de tipos de datos de interés para el dominio del problema. Este proceso (Figura 1) consta de 5 fases: preparación de datos (Sección 2.1), definición de vecindarios (Sección 2.2), modelado gráfico de relaciones espaciales (Sección 2.3), minado de subgrafos frecuentes (Sección 2.4), y evaluación de resultados (Sección 2.5).

2.1. Preparación de los datos

La primera etapa del proceso de explotación de información consiste en la preparación de los datos espacialmente referenciados. Estos datos, disponibles en distintas fuentes, deben ser codificados de tal forma que las relaciones entre los mismos puedan ser extraídas en etapas posteriores.

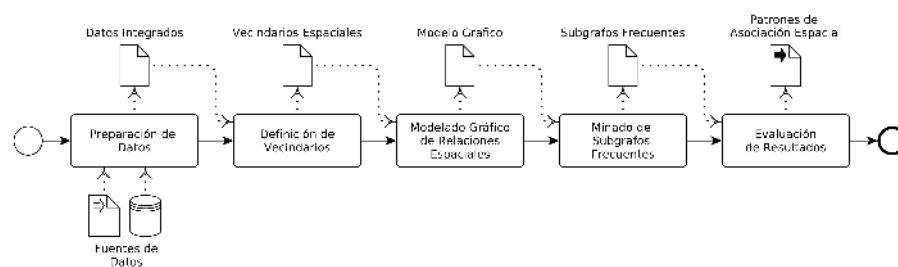


Figura 1. Proceso de Explotación de Información para el descubrimiento de Asociaciones Espaciales.

Dos conjuntos de datos deben ser determinados: los datos objetivo, alrededor de los cuales se desea encontrar patrones de conocimiento, y los datos relevantes, que se relacionarán con los datos objetivos y proveerán información útil para el dominio del problema [23]. Estos conjuntos deberán ser limpiados y preparados, resolviendo errores, inconsistencias, datos nulos y anomalías. Además, nuevos atributos derivados podrán ser generados en caso de que se lo requiera, o incluso nuevos datos espaciales a partir de los disponibles en las distintas fuentes de datos de entrada.

Esta etapa requiere un esfuerzo considerable y posiblemente varias iteraciones, por lo que es recomendable implementar todo el proceso mediante una metodología apropiada, tal como CRISP-DM [24] o MoProPEI [25].

2.2. Definición de Vecindarios

Como se ha mencionado anteriormente, los patrones encontrados en el conjunto de datos pueden poseer una alta probabilidad de ocurrencia en solo algunos lugares del espacio contemplado, por lo que una búsqueda local o regional debe ser realizada [7].

Para esto se propone la definición de particiones del conjunto de datos, a las que denominaremos *vecindarios*, y la ejecución de la búsqueda de asociaciones dentro de cada una de ellas. Estos vecindarios pueden ser definidos previamente, correspondiéndose a características del dominio del problema, o bien pueden ser definidos mediante técnicas de segmentación o *clustering* espacial.

En el presente trabajo se recomienda la utilización de algoritmos de agrupamiento o bien basados en densidad o bien basados en distancia sobre los atributos espaciales. Esta preferencia está fundamentada en la Primera Ley de la Geografía, la cual establece que existe una mayor relación entre los objetos espaciales que se encuentran cercanos entre sí [26–29]. Opciones a ser utilizadas en la implementación del proceso son los algoritmos basados en densidad de la familia DBSCAN [30–35].

2.3. Modelado Gráfico de Relaciones Espaciales

Una vez generados los vecindarios espaciales es necesario explicitar por cada uno de ellos las relaciones espaciales entre los objetos espaciales objetivo y los objetos de referencia. En función del dominio del problema distintos tipos de relaciones pueden ser consideradas [6]: (i) Relaciones Geométricas, si están basadas en la distancia euclidiana; (ii) Direccionales, basadas usualmente en el concepto de vecindario en grillas, utilizándose el modelo de 8 direcciones (Norte, Noreste, Este, Sudeste, Sur, Suroeste, Oeste, Noroeste); (iii) Topológicas, independientes del concepto de distancia y aplicables a cualquier tipo de objeto espacial, se utiliza normalmente el modelo de 9-intersecciones [36–38]; (iv) Híbridas, si responden a múltiples categorías anteriores, como el paralelismo y la perpendicularidad.

Para este modelado se hace uso de la teoría de grafos por su estrecha relación con los predicados y el tipo de patrón que se busca [11] (Figura 2). Sea entonces

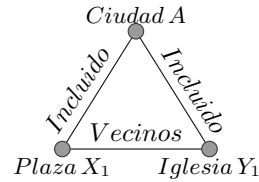


Figura 2. Ejemplo simplificado del modelado de relaciones espaciales mediante multigrafos. Se puede apreciar una instancia del patrón de asociación mencionado en la sección 2.

un multigrafo sin bucles $G = (V, E, L, K, \phi, P_l, P_k)$ en donde V es un conjunto de vértices que se corresponden con las instancias de los objetos espaciales, E es un conjunto de aristas que se corresponden con las relaciones entre los objetos espaciales, L es un conjunto de características o valores de los atributos de los objetos espaciales, K es un conjunto de características o atributos de las relaciones espaciales, $\phi : E \rightarrow \{\{v_1, v_2\} | v_1 \in V \wedge v_2 \in V \wedge v_1 \neq v_2\}$ es la función de incidencia de las relaciones espaciales sobre los objetos espaciales, $P_l \subseteq V \times L$ es la relación de etiquetado de características de los objetos espaciales, y $P_k \subseteq E \times K$ es la relación de etiquetado de características de las relaciones espaciales.

De esta manera, al poseer múltiples aristas por cada par de vértices, se pueden modelar distintos tipos de relaciones, que a su vez pueden ser etiquetadas con diversos atributos y, de la misma manera, varios atributos de los objetos espaciales pueden ser tenidos en cuenta. Por otro lado, si la situación lo amerita, estas relaciones pueden ser explicitadas mediante el uso de lógica difusa [12].

2.4. Minado de Subgrafos Frecuentes

Con el propósito de extraer las asociaciones en el espacio con alta tasa de ocurrencia se propone la búsqueda de subgrafos frecuentes dentro de los grafos asociados a cada vecindario espacial. Dado un multigrafo $G = (V, E, L, K, \phi, P_l, P_k)$ como el descrito en la sección anterior, el problema de la búsqueda de subgrafos frecuentes en un único gran grafo consiste en encontrar subgrafos $G_i \subset G$ que sean recurrentes en G o, dicho de otra manera, que se repitan múltiples veces dentro del grafo original (Figura 3). Es necesario remarcar que para que un grafo sea isomorfo con otro, las etiquetas de los nodos y las aristas también deben coincidir.

Estos subgrafos frecuentes representan aquellas relaciones entre tipos de objetos espaciales que poseen una alta tasa de ocurrencia dentro del modelo gráfico de los objetos espaciales, correspondiendo así a asociaciones espaciales. Al poseer múltiples relaciones y etiquetas, diversas situaciones pueden ser exploradas simultáneamente.

Múltiples algoritmos han sido diseñados para la búsqueda de subgrafos frecuentes en grandes grafos, teniendo además diversas formas de medir la relevan-

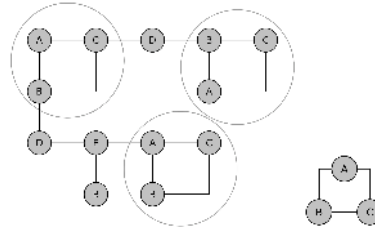


Figura 3. Ejemplo de subgrafo frecuente en un grafo simple sin etiquetas en las aristas.

cia de cada subgrafo frecuente en función de parámetros como la recurrencia o el índice de compresión de información. Algunos ejemplos de algoritmos bien conocidos son IncGM+, FSSG, SUBDUE, HSIGRAM, GREW, entre otros [39–41]

Como resultado de esta fase se obtiene un conjunto de subgrafos frecuentes por cada vecindario, que deberán ser analizados con el fin de obtener conocimiento útil para la toma de decisiones.

2.5. Análisis de Resultados

Como paso final resta el filtrar aquellos subgrafos frecuentes triviales (aquellos no novedosos en el dominio del problema) y traducir los subgrafos frecuentes restantes a predicados compuestos a fin de poder ser utilizados por la inteligencia de negocios.

Este paso puede realizarse de manera automatizada o bien ser realizado por el analista con conocimiento en el dominio del problema o con la ayuda de un experto.

3. Prueba de Concepto

La prueba de concepto efectuada en esta sección pretende ilustrar el funcionamiento del proceso de explotación de información propuesto mediante el uso de distintas herramientas de programación y minería de datos para su implementación.

Como entrada del proceso se dispone de 10 archivos con la ubicación de distintos tipos de edificios de la Ciudad Autónoma de Buenos Aires y alrededores (Bibliotecas, Clínicas, Oficinas de Correo, Instalación Deportiva, Discotecas, Escuelas, Estaciones de Servicio, Iglesias, Museos y Estaciones de Policía). Por cada uno de ellos, en la etapa de preparación de los datos, se generaron mediante la herramienta QGis¹ objetos espaciales puntuales que consisten en dos atributos espaciales con información sobre la Latitud y la Longitud, y un atributo no espacial con el tipo de edificio. Posteriormente se seleccionaron solo aquellos que se encuentran dentro de la Ciudad Autónoma de Buenos Aires para acotar el dominio de búsqueda, resultando 742 puntos (Figura 4, Izquierda).

¹ <http://qgis.org/> [10 de julio de 2018]

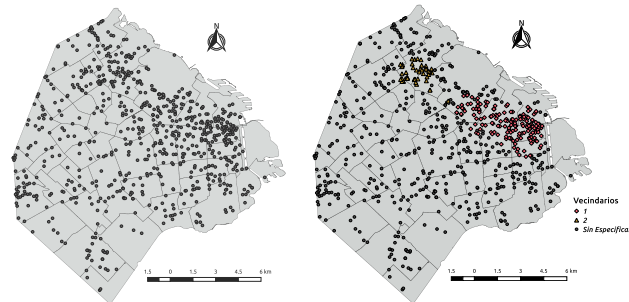


Figura 4. Objetos espaciales generados para la prueba de concepto (Izquierda). Vecindarios generados para la prueba de concepto utilizando HDBSCAN (Derecha).

Posteriormente, en la fase de definición de vecindarios, se utilizó el algoritmo HDDBSCAN [35] de la librería 'dbscan'² de R sobre los atributos espaciales de los datos, generándose dos vecindarios con un mínimo de 50 puntos en cada uno (Figura 4, Derecha), para reducir el número de vecindarios generados por fines ilustrativos.

Luego, por cada uno de los vecindarios se explicitaron las relaciones geométricas de vecindad entre los objetos espaciales mediante un grafo, generándose un vértice por cada uno de ellos etiquetado con el tipo de edificio al que corresponde, y un arista etiquetado con la etiqueta "Cerca" si la distancia entre los vértices adyacentes resulta menor a 150 metros. Como resultado se obtuvo un grafo con 71 vértices y 45 aristas para el vecindario 1, y un grafo con 15 vértices y 11 aristas para el vecindario 2.

Para la obtención de los subgrafos frecuentes en cada uno de los vecindarios se utilizó el algoritmo SUBDUE implementado en la aplicación Subdue Graph Miner³, obteniéndose para el vecindario 1 un subgrafo con una tasa de compresión del 15,5% del que se extrajo el predicado $Correo(X) \wedge Discoteca(Y) \wedge Cerca(X, Y)$, y dos subgrafos del vecindario 2, con una tasa de compresión del 27,2%, de los que se extrajeron los predicados $Clinica(X) \wedge Correo(Y) \wedge Cerca(X, Y)$ y $Correo(X) \wedge Deporte(Y) \wedge Cerca(X, Y)$.

Como se puede observar, el patrón observado en el primer vecindario no se evidencia en el vecindario número 2, y viceversa, por lo que son patrones que solo poseen una alta probabilidad de ocurrencia local.

4. Conclusión

En el presente trabajo se presenta el diseño de un proceso de explotación de información para el descubrimiento de patrones de asociación en datos espacialmente referenciados. Este proceso resulta suficientemente flexible para poder

² <https://cran.r-project.org/web/packages/dbscan/dbscan.pdf> [10 de julio de 2018]

³ <http://ailab.wsu.edu/subdue/> [10 de julio de 2018]

contemplar multiplicidad y variedad de relaciones espaciales extraídas de diferentes tipos de datos espaciales, haciendo uso de estructuras de grafo para el modelado de las mismas. Se tiene en cuenta además los fenómenos de heterogeneidad y autocorrelación espacial definiendo vecindarios en los cuales se realizará la búsqueda de este tipo de regularidad.

Por otro lado, se presenta una prueba de concepto utilizando datos reales en la cual se pone en evidencia como se realiza la implementación del proceso utilizando distintas herramientas y técnicas para cada una de las etapas propuestas.

En futuros trabajos se hará énfasis en estrategias de implementación de cada una de las fases según el dominio del problema a fin de reducir los tiempos computacionales de ejecución de las actividades para grandes volúmenes de datos con gran cantidad de relaciones entre los mismos. Además se evaluarán distintas estrategias de implementación de extracción de relaciones difusas.

Agradecimientos

La investigación presentada en este artículo fué parcialmente financiada por el Programa de Doctorado para el Fortalecimiento de Áreas I+D+i (2016-2020) de la Universidad Tecnológica Nacional y por el Proyecto de Investigación 80020160400001LA de la Universidad Nacional de Lanús.

Referencias

1. Mennis, J., & Guo, D. "Spatial data mining and geographic knowledge discovery—An introduction." *Computers, Environment and Urban Systems* 33.6 (2009): 403-408.
2. Koperski, Krzysztof, and Jiawei Han. "Discovery of spatial association rules in geographic information databases." *International Symposium on Spatial Databases*. Springer, Berlin, Heidelberg, (1995).
3. Rottoli, Giovanni Daián. "Formulación de Procesos para una Ingeniería de Explotación de Información Espacial." *XXI Ibero-American Conference on Software Engineering*. Doctoral Symposium. Bogotá, Colombia, (2018).
4. Leung, Yee. "Knowledge Discovery in Spatial Data." Ed. Manfred M. Fischer et al. 1st ed. Vol. 45. Hong Kong: Springer, (2011).
5. Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases(VLDB)*. Vol. 1215. (1994).
6. Appice, Annalisa et al. "Discovery of Spatial Association Rules in Geo-Referenced Census Data : A Relational Mining Approach." *Intelligent Data Analysis* 7 (2003): 541-566.
7. Simpson, Edward H. "The interpretation of interaction in contingency tables." *Journal of the Royal Statistical Society. Series B (Methodological)* (1951): 238-241.
8. Saleb, Ansaf, and Christel Vrain. "An Application of Association Rules Discovery to Geographic Information Systems." *European Conference on Principles of Data Mining and Knowledge Discovery*. (2000). 613-618.

9. Sutjipto, Suci Sri Utami, Imas Sukaesih Sitanggang, and Baba Barus. "Potential Usage Estimation of Ground Water using Spatial Association Rule Mining." *Telecommunication Computing Electronics and Control (TELKOMNIKA)* 15.1 (2017): 504-511.
10. Malerba, Donato, and Francesca A Lisi. "Discovering Associations between Spatial Objects : An ILP Application." *International Conference on Inductive Logic Programming* (2001): 156-163.
11. Malerba, Donato, and Floriana Esposito. "Mining Spatial Association Rules in Census Data." *An International journal for research in official statistics (ROS)* 5.1 (2002): 19.
12. Ladner, Roy, Frederick E Petry, and Maria A Cobb. "Fuzzy Set Approaches to Spatial Data Mining of Association Rules." *Transactions in GIS* 7.1 (2003): 123-138.
13. Bai, Hexiang et al. "A Method for Extracting Rules from Spatial Data Based on Rough Fuzzy Sets." *Knowledge-Based Systems* 57 (2014): 28-40.
14. Zhang, Anshu. "Uncertainty-Based Spatial Association Rule Mining." The Hong Kong Polytechnic University, 2017.
15. Goudarzi, Amir Hossein, and Nasser Ghadiri. "A Hybrid Spatial Data Mining Approach Based on Fuzzy Topological Relations and MOSES Evolutionary Algorithm." *arXiv preprint arXiv:1704.06621* 2017: 1-25.
16. Lee, Ickjai. "Mining Multivariate Associations within GIS Environments." *Innovations in Applied Artificial Intelligence* 3029 (2004).
17. Yang, Hui, Srinivasan Parthasarathy, and Sameep Mehta. "Mining Spatial Object Associations for Scientific Data." *IJCAI International Joint Conference on Artificial Intelligence II* (2005): 902-907.
18. Ding, Wei et al. "A Framework for Regional Association Rule Mining in Spatial Datasets." *Data Mining, 2006. ICDM'06. Sixth International Conference on.* (2006). 851-856.
19. Ding, Wei et al. "A Framework for Regional Association Rule Mining and Scoping in Spatial Datasets." *GeoInformatica* 15.1 (2011): 1-28.
20. Sha, Zongyao, and Xiaolei Li. "Mining Local Association Patterns from Spatial Dataset." *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (2010): 1455-1459.
21. Sha, Z., X. Tan, and Y. Bai. "Localized Spatial Association: A Case Study for Understanding Vegetation Successions in a Typical Grassland Ecosystem." *Communications in Computer and Information Science* 482 (2015): 33-45.
22. Dao, Thi Hong Diep, and Jean Claude Thill. "The SpatialARMED Framework: Handling Complex Spatial Components in Spatial Association Rule Mining." *Geographical Analysis* 48.3 (2016): 248-274. DOI: 10.1111/gean.12094
23. Bogorny, Vania, Paulo Martins Engel, and Luis Otavio Alvares. "GeoARM: An Interoperable Framework to Improve Geographic Data Preprocessing and Spatial Association Rule Mining." *18th International Conference on Software Engineering and Knowledge Engineering (SEKE)* (2006): 79-84.
24. Wirth, Rüdiger, and Jochen 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.* (2000).
25. Martins, Sebastian, Patricia Pesado, and Ramón García-Martínez. "Information Mining Projects Management Process." *28th International Conference on Software Engineering & Knowledge Engineering (SEKE)*. (2016): 504-509. ISBN 1-891706-39-X.
26. Tobler, Waldo R. "Cellular geography." *Philosophy in geography.* Springer, Dordrecht. (1979): 379-386.

27. Miller, Harvey J. "Tobler's first law and spatial analysis." *Annals of the Association of American Geographers* 94.2 (2004): 284-289.
28. Tobler, Waldo. "On the first law of geography: A reply." *Annals of the Association of American Geographers* 94.2 (2004): 304-310.
29. Duan, Jiangli, Lizhen Wang, and Xin Hu. "The effect of spatial autocorrelation on spatial co-location pattern mining." *Computer, Information and Telecommunication Systems (CITS), 2017 International Conference on*. IEEE. (2017).
30. Sander, Jörg, et al. "Density-based clustering in spatial databases: The algorithm gdbscan and its applications." *Data mining and knowledge discovery* 2.2 (1998): 169-194.
31. Smiti, Abir, and Zied Elouedi. "Fuzzy density based clustering method: Soft DBSCAN-GM." *Intelligent Systems (IS), 2016 IEEE 8th International Conference on*. IEEE. (2016).
32. Birant, Derya, and Alp Kut. "ST-DBSCAN: An algorithm for clustering spatial-temporal data." *Data & Knowledge Engineering* 60.1 (2007): 208-221.
33. Zhu, Ye, Kai Ming Ting, and Mark J. Carman. "Density-ratio based clustering for discovering clusters with varying densities." *Pattern Recognition* 60 (2016): 983-997.
34. Sharma, Arvind, R. K. Gupta, and Akhilesh Tiwari. "Improved Density Based Spatial Clustering of Applications of Noise Clustering Algorithm for Knowledge Discovery in Spatial Data." *Mathematical Problems in Engineering* 2016 (2016).
35. Campello, Ricardo JGB, et al. "Hierarchical density estimates for data clustering, visualization, and outlier detection." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10.1 (2015): 5.
36. Egenhofer, Max J., and Robert D. Franzosa. "Point-set topological spatial relations." *International Journal of Geographical Information System* 5.2 (1991): 161-174.
37. Egenhofer, Max J., and John Herring. "Categorizing binary topological relations between regions, lines, and points in geographic databases." *The* 9.94-1 (1990): 76.
38. Egenhofer, Max J., Eliseo Clementini, and Paolino Di Felice. "Topological relations between regions with holes." *International Journal of Geographical Information Science* 8.2 (1994): 129-142.
39. Abdelhamid, Ehab, et al. "Incremental Frequent Subgraph Mining on Large Evolving Graphs." *IEEE Transactions on Knowledge and Data Engineering* 29.12 (2017): 2710-2723.
40. Kavitha, D., V. Kamakshi Prasad, and J. V. R. Murthy. "Finding Frequent Subgraphs in a Single Graph based on Symmetry." *International Journal of Computer Applications* 146.11 (2016).
41. Bianco, Santiago. "Análisis Comparativo de Algoritmos de Minería de Subgrafos Frecuentes." *Revista Latinoamericana de Ingeniería de Software* 4.2 (2016): 111-142.