HAL open science

Join indices as a tool for spatial data mining

Karine Zeitouni, Laurent Yeh, Marie-Aude Aufaure

▶ **To cite this version:**

Karine Zeitouni, Laurent Yeh, Marie-Aude Aufaure. Join indices as a tool for spatial data mining. International Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining, TSDM 2000, Sep 2000, Lyon, France. pp.105-116, 10.1007/3-540-45244-3_9 . hal-04371588

HAL Id: hal-04371588
https://hal.univ-lille.fr/hal-04371588

Submitted on 3 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Join indices as a tool for spatial data mining

Karine Zeitouni[1] , Laurent Yeh[1] , Marie-Aude Aufaure[2]

[1] Prism Laboratory - University of Versailles
45, avenue des Etats-Unis - F-78 035 Versailles Cedex
{last name.first name}@prism.uvsq.fr
[2] LISI Laboratory - INSA of LYON bat. 501
F-69 621 Villeurbanne Cedex
Marie-Aude.Aufaure@lisi.insa-lyon.fr

Abstract. The growing production of maps is generating huge volume of data stored in large spatial databases. This huge volume of data exceeds the human analysis capabilities. Spatial data mining methods, derived from data mining methods, allow the extraction of knowledge from these large spatial databases, taking into account the essential notion of spatial dependency. This paper focuses on this specificity of spatial data mining by showing the suitability of join indices to this context. It describes the join index structure and shows how it could be used as a tool for spatial data mining. Thus, this solution brings spatial criteria support to non-spatial information systems.

1 Introduction

In the area of Knowledge Discovery in Databases (KDD), Spatial Data Mining (**SDM**) is becoming an important issue that offers new prospects for many data analysis applications. Spatial data mining performs data driven analysis of large size spatial databases. The propensity of such large databases is due to the easiness of acquiring geodata. Indeed, geodata sets become available via Internet and the development of geo-coding tools allow attaching spatial data to a-spatial database according to address fields.

Spatial data mining tasks are considered as an extension of Data Mining (**DM**) tasks [11] in which spatial data and criteria are combined. As According to [15], these tasks aim at: (i) summarising data, (ii) finding classification rules, (iii) making clusters of similar objects, (iv) finding associations and dependencies to characterise data, and (v) looking for general trends and detecting deviation. They are carried out using specific methods, some of which are derived from statistics and others from the field of machine learning. As shown in [35], both approaches converge in the way they use spatial relationships. The problem is that the implementation of all those methods is not straightforward and right now, there is a lack of spatial data mining tools.

The goal of this paper is to introduce the use of the well-known “join index” structure as a tool for spatial data mining. Indeed, we point out that one major difference between traditional data mining and spatial data mining is the notion of spatial relationships. These spatial relationships correspond to the spatial join operator and could be resolved by using “join index”. We outline that this join index is a powerful tool to transform traditional database queries into spatial ones. A basic relational table is thereby sufficient to represent this join index. This has the great advantage to reduce spatial data mining to conventional data mining and provides an easy and efficient support of spatial data mining.

The rest of the paper is organized as follows: In the next section, we briefly introduce related works in the area of spatial data mining. In section 3 we emphasize the particularity and the importance of spatial relationships. These spatial relationships are then materialized by join indices in section 4. In this section, we outline the similarities between “join index” and “contiguity matrix” used in the field of spatial statistics. Section 5 discusses technical aspects in their implementation. We conclude in the last section.

2 Spatial databases and data mining

This section briefly introduces spatial databases, and then related works in the fields of spatial statistics and spatial databases are described.

2.1 Spatial databases context

Spatial database systems are a great part of Geographical Information Systems (GIS) [22, 21, 23]. They store and manage huge volume of geographical entities such as road sections or lakes. Each entity combines the location description and other a-spatial data related to the entity such as the lake name or capacity.

A spatial database is organized in a set of thematic layers. A thematic layer is a collection of geographical objects that share the same structure and properties. A theme can represent a road network, and another can represent towns. This allows to selectively use the relevant themes for a specific purpose.

2.2 Advances in spatial data mining

The aim of Spatial Data Mining [29] is to extract knowledge, spatial interactions and other properties that are not explicitly stored in the database. This process inherits from traditional data mining which could be performed by “a set of tools that allow to extract automatically or semi-automatically interesting and understandable knowledge (rule, regularities, patterns, associations..) from a database” [11].

The specificity of SDM lies in its interaction in space. In effect, a geographical database constitutes a spatio-temporal continuum in which properties concerning a particular place are generally linked and explained in terms of the properties of its neighbourhood. We can thus see the great importance of spatial relationships in the analysis process. Temporal aspects for spatial data are also a central point but are rarely taken into account.

Since traditional data mining methods [11] do not support location data or the implicit relationships between objects, it is necessary to develop specific methods for spatial data mining. As it is well known, geometric data and processing are more complex than traditional ones. Spatial applications also generate a huge volume of data. For these reasons, calculating these spatial relationships is time consuming. One major problem is to optimise analysis methods taking into account the huge volume of data and the complexity of spatial relationships.

Research in the field of SDM comes from both spatial statistics and databases. There is an abundant bibliography about spatial analysis [12, 13] existing before the GIS era. Many research works have been done to measure global and local spatial auto-correlation [7]. In the field of geostatistics [26], *kriging* technique analyses spatial trends. Recent research in interactive techniques for Exploratory Spatial Data Analysis (ESDA) have been developed [2, 1, 19, 20]. Openshaw [28] has developed a prototype using parallel computing to identify clusters. Some research works have also been done to extend the multivariate statistical analysis in order to support contiguity constraints [5]. From our point of view, spatial statistics methods are part of spatial data mining, since they provide data-driven analysis. Some of those methods are now implemented in operational GIS or analysis tools.

In the field of databases, spatial data mining algorithms have been proposed and prototypes have been developed. GeoMiner [14, 15] is an extension of DBMiner that integrates OLAP techniques and a coupling between a GIS and a database. Most of the proposed algorithms in GeoMiner are based on a priori knowledge: the concept hierarchies. Ester at al. devised a structure-of-neighbourhood graph[10], on which some algorithms are based. They have also worked on SDM methods such as clustering method (extension of DBSCAN with an R*Tree), classification (extension of ID3 and DBLearn), association rules (based upon an efficient spatial join), characterisation and spatial trends [10]. STING (University of California) uses a hierarchical grid to perform optimisation on the clustering algorithm [33]. We might also mention work on data warehousing dedicated to spatial data (University of Laval) [3].

In conclusion to this brief state of the art, we want to point out the similarities of the spatial statistics approach and the database one. The main similarity is the use and the importance of neighbourhood relationships. Contiguity matrices are used to compute spatial auto-correlation and neighbourhood graphs represent a secondary structure useful for many SDM algorithms. Another similarity is the current use of distance criteria in defining neighbouring objects as well as in clustering methods.

3 Spatial relationships

Spatial relationships represent an essential characteristic in real world. They show spatial influences between entities. Indeed, observations located near to one another in space tend to share similar attribute values. This is known as one of the "1st law in geography" specified by Tobler [31]. Moran has defined a measurement of spatial auto-correlation between nearby data since 1948 [7]. Anselin has refined this in local auto-correlation indices [2] qualifying the correlation of each entity value with the values of its neighborhood.

Spatial and local auto-correlation only consider the interactions within one theme. In reality, thematic layers are often strongly correlated. For instance, precipitation and population density maps are correlated: the population depends on the agricultural production that depends on the precipitation. For this reason, we distinguish two kinds of relationships. On the one hand, those that link objects of the same thematic layers (we call it intra-theme). On the other hand, those involving two different layers (called inter-themes), like the inclusion of an accident location within a road section or within a county.

Spatial relationships have also been formalized in spatial database theory and extended to many topological relationships [8] such as intersection and inclusion. Moreover, metric relationships, e.g. using distance criteria, are also admitted as spatial relationships.

Spatial data mining methods make intensive use of spatial relationships. That actually distinguishes these methods from conventional data mining methods. This shows the main role of such relationships in spatial data analysis and mining. However, this uncovers many problems and specific properties as explained below.

First, those relationships are usually implicit except in topological models [4]. Their resolution leads to spatial join that is known as a complex and costly operation in spatial database. Therefore, efficient support of spatial joins should be studied.

Second issue is to allow specifying those relationships as spatial knowledge and to use it in the analysis. Indeed, some spatial relationships are pre-established and constitute spatial integrity constraints. For instance, car crash is constrained to be located on a road section. Hence, methods such as clustering should be modified to consider a linear network instead of an open space.

Third, inclusion spatial relationships, and more generally, the difference in the sizes or the weights of spatial units [23, 16], make it difficult to assess quantitative data. This involves the extension of those relationships to weighted relationships or more generally using a complex model including behavioral functions.

The last two points require the spatial relationships resolution, which is precisely the first point. This point is mainly the focus of our current work.

4 Spatial join index

As contiguity matrix is an important component in spatial statistics, we show in this section that join index is also an important component for spatial data mining.

First, we describe the idea of a join index, then, discuss its different extensions to spatial databases. Section 4.3 shows its similarity with contiguity matrix allowing its direct utilisation in spatial statistics, such as auto-correlation computation, as well as for spatial data mining. Section 4.4 exposes another advantage of spatial join index, using it to simplify spatial analysis by reducing it to relational case. This is possible thanks to the very simple representation of a spatial join index by a relational table.

4.1. Join indices

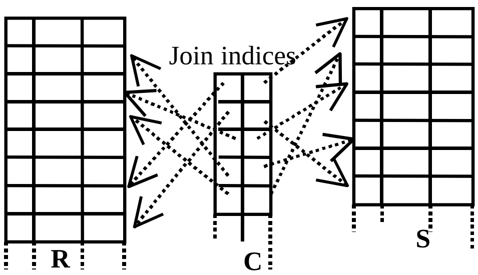

Figure 1: Join index.

Obj1	Obj2	Distance
R-1	S-9	2.34
R-1	S-1	3.45
R-2	S-11	7.23
R-2	S-13	3.22
R-2	S-14	5.34
R-3	S-18	3.44
R-3	S-16	3.68
…		

Figure 2 : Spatial filter join index

The join-index method has been proposed by [32] as a valuable technique for increasing speed of query evaluation in Data Base Management System (DBMS) [23]. It constitutes a table that stores indice pairs (called join indices), each referencing an object of each input object collection (c.f. figure 2). Figure 1 shows this structure where C is the join index an **R** and **S** the two object collections. Here **R-1** to **R-n** are object identifiers [18]. The pairs of indices refer to objects that match the join criterion. In this figure, the join indices represent the distance between two objects (representing entities). In spatial topological relationships like inclusion, intersection, the third column could disappear because the distance is zero and does not change.

4.2. Extensions to spatial join indices

Conversely to relational joins, spatial joins use as many criteria as possible spatial relationships. There are at least three alternatives to extend join indices to spatial databases.

- The first is a direct application, leading to build one join index set for each spatial predicate. That means exact computation of all these predicates, which is time and storage space consuming.

- A second way is to build only one join index set by adding columns for each kind of spatial relationship as in [10]. This improves a little bit the storage volume but the computing time is still high.

  The two previous solutions are not sufficient to deal with distance based joins when the distance is specified dynamically.

- The third solution proposes to store a coarse computation of the spatial criteria rather than the exact spatial join criteria. This is allowed by the definition of one join index having a third column for a distance value of the referenced spatial objects [30]. This structure (see figure 2) will play the role of a filter for most spatial joins. Indeed, the topological relationships as well as the metric ones can be deduced from using a simple selection in this join index. Notice that in this solution as well as for any distance based criteria, one usual optimisation of storage space is to limit the distance to a given maximal distance (called scope). Indeed, instead of storing all combinations of couples of object references, only those having a reasonable distance (under the scope) will be retained. Another alternative for optimising distance based join indices is described in [34].

4.3. Homogeneity with contiguity matrix

Besides, in spatial analysis area, as for the computation of Moran's and Geary's indices [7], it is usual to use what is called "contiguity" matrix. This matrix is defined as M where M(i,j) = 1 when the two objects i and j are contiguous, and M(i,j) = 0 otherwise. As this is a sparse matrix, it is usual to represent it in a more compact format as an array structure. Actually, it is the same structure than the above join index by using as object identifiers the row numbers. This homogeneity means that the same structure could serve in the new spatial data mining methods and the statistical spatial analysis methods. The second interest is to provide a generic way to yield contiguity matrices under any criteria.

Notice that contiguity matrices are square matrices defined on a unique data set (i.e. intra-theme) while join indices could apply to compare two data sets (i.e. inter-themes). So, spatial join indices are more general than contiguity matrices.

4.4. Reducing to relational DM

After the computation step of join indices, using them has three main advantages that we outline in this sub-section. The first advantage is to be able to provide spatial analysis functionalities to a system that do not have this capability. The second one is to allow the use of powerful query language. The third one is that such a structure is used as an accelerator.

For the first advantage, join indices could be implemented using an external process to the given system. With modern system capabilities as DLL (Dynamic Link Library) or IPC (Inter-Process Communications) and Internet capabilities (using Socket), it is easy for most widespread tools to raise data analysis in spatial domain. This means a way for spatial data mining. This extensibility is described with technical aspects in the next section.

For the second advantage, as a join index has a tabular representation in relational DBMS, it could be handled in the same way as other tables. This allows using all the power of database query languages such as the normalized SQL language. For instance, by selecting and matching entities according to their attributes, one can focus the analysis on relevant data.

For the third advantage, many optimising techniques in spatial database domain have aimed at increasing the performance of spatial data retrieval including spatial joins. One category uses spatial access methods (e.g. R-tree, Grid-file, etc.). Here, using join indices is another alternative, based on the principle of pre-computing join operations. The interest in pre-computing is to do it once and to use it many times allowing an important gain while producing the final join result. Indeed, for the analysis purpose, interactive hypotheses tests become possible.

Performance measures have shown the gain in terms of processing time by comparing execution time with and without the pre-computed index (see figure 3).

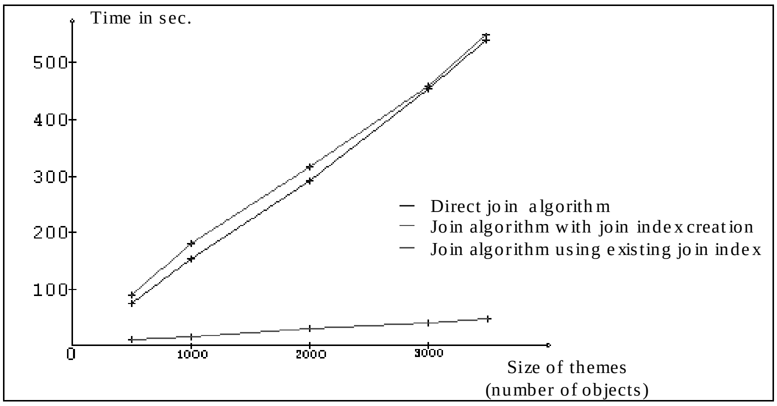

Figure 3: Performance measurement in using spatial join index

In many applications, the construction of join index structures could be done while loading data in the system. This happens in reality very few times. In addition, this computation could be done offline.

In summary, due to its tabular representation, join index is a pragmatic and efficient tool for spatial data mining.

4.5. Join index use in SDM

There exist two possibilities in using spatial join indices for SDM purpose:

1. during the stage of data selection preparing them to SDM tasks: as it is well known, previously to DM, KDD process includes a phase that transforms the initial data to a target data set. This data set holds relevant information for the analysis and should have the expected format of DM algorithms. The most used format is one table. This table is generally built by joining many initial database tables holding the relevant attributes. This corresponds to de-normalization in relational databases. Following this process, a pragmatic approach to spatial KDD is to, first, spatially join relevant thematic layers and then, apply DM tasks.
2. by strongly integrating them in specific SDM algorithms [9]. Here, join indices are viewed as a low-level library only manipulated internally. This is more complex to implement and is less portable to existing systems. However, strong integration achieves better performance in term of processing time.

5. Spatial join index implementation

This section discusses technical aspects of spatial join index for a user who wants to implement them within an existing DBMS. It shows, first, the use of join indices at the user level. Then, it raises the problem of object referencing and studies diverse solutions. At the representation level, two possibilities are described and compared. Then an implementation approach is given.

5.1. How does it appear for users ?

For users, join indices appear as normal tables of the database schema. So they can be manipulated using SQL. For instance, in the query of figure 4, where each of R and S contain a geometric attribute and the join index `Indices_dist_RS` stores all distances between R and S, this spatial join is replaced by two joins (figure 4, lines (1) and (2)). In this case, relation databases are sufficient, once join indices have been computed.

```
Select R.name, R.Population

From R, S, Indices_dist_RS I
Where I.obj1 = R.id        (1)
And    I.obj2 = S.id       (2)
And    I.distance <  50
```

Figure 4: Use of join indices in a query.

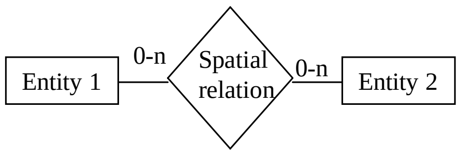

Figure 5: Spatial relation as a semantic link between entities.

This is a simple way to consider spatial relation evaluation. Indeed, this works as if all spatial relationships were semantic relations (in the well-known Entity/Relation data model [25]) and defined on each combination of spatial entities. Relation (diamond-shaped on figure 5) traduces a semantic relationship between two entities (rectangular in figure 5). This diagram shows the conceptual level. It is implemented in relational data model by tables. Thus, join indices tables are the physical representation of spatial relations. This simplifies the concepts for users by replacing complex spatial join concept by most familiar table joins.

Indeed, the main difference is that join indices are automatically deduced by a program instead of being a priori entered information. It is unrealistic to define conceptual model with all the above combinations of links, because it is a combinatorial problem. Only the useful combination could be pre-computed by the database designer.

5.2. Object reference choice

One main difficulty in using join indices, is to preserve a unique reference to a spatial value (similar to Object Identifier). Very few relational DBMS provide at the user’s level an Object Identifier [18] even if in Object DBMS, there is the notion of Object Identifier. In relational DBMS, the closest notion is the key attributes. However, the key could change by updating. Moreover, as a result of a query, we could lose the key attribute. Another possibility is to use the object row number in the table. This corresponds to the formal definition of contiguity matrices and systems like ArcView and Oracle allow handling rows. But, row numbers are no more available when objects are deleted or inserted. For these reasons, the use of join indices at the user’s level in relational DBMS could be tedious. Integrity constraints or triggers could be implemented to prevent the key change. Fortunately, spatial databases are rather static.

5.3. One or several join index sets between entities

Between two object sets, many spatial relationships (inclusion, neighborhood…) could be interesting. The problem is how to manage all these relationships in join indices [26]. Two approaches could be used to represent these data. The first approach is to store all the couples of indices that represent one relationship between two entities in individual sets. Then, we could name each set as the relationship. In the second approach (figure 6), we use a flag for the existence of a relationship between two entities. Thus, in one record, we could set several relationships. As illustrated in figure 6, the two first values are references to respective objects. The value '1' means the existence of a given relationship and null value for none relationship.

….
<s1, s2, null, null, 1, null, 1,…>
<s1, s3, 1, null, 1, null, null,…>
…

Figure 6: Example of indices with multiple relationships.

	Advantage	**Disadvantage**
Using several join indices (**SJI**) sets	Each set uses the requisite memory	Less efficient for multi-criteria queries
Using one join index (**OJI**) set	Fast to process	Many empty values (few objects match for a given relationship).

Figure 7: One set versus several sets.

Figure 7 compares the advantages and the disadvantages of each approach. Using several join indices (SJI) sets is optimal in the set size. Indeed, each set holds only a given spatial relationship. Each set stores only the couple of indices that match a given spatial relationship. However, when a query uses several spatial criteria (e.g. forest close to lakes AND included in Versailles City), it involves several join processing (one for each spatial relationship and one to compute the final result). Then, for SJI, the storage is more efficient but the processing cost could be important because of join operators and the query expression could be more complex.

On the contrary, using one join index (OJI) requires only two join operators for each pair of sets involving one or several (multi-criteria) spatial relationships within one query. However, the size of the set could be huge in case it integrates distance relationship. Indeed, distance involves all the couples of entities. The main consuming space is due to the two object identifiers. To overcome this size problem, a solution could be to represent in one bit a topological relationship. Moreover, we could represent an exclusive data between distance and a set of topological relationships. This needs one more bit to switch between these two representations.

5.4. Extensibility of existing tools

Join index provides a pragmatic solution for existing tools and DBMSs that do not handle spatial features or for GISs that do not support some spatial join criteria. For instance, in ArcView, distance based spatial joins are not supported. An external program could compute it and it appears as a simple table that could be integrated in the system.

We have applied this approach in our server prototype. The `Arcview` system sends a signal to server for computing a set of join indices. Then the result is stored as a `dbf` data file (directly readable in `Arcview`), and the server sends a signal to `ArcView` for synchronizing the next step of processing. This simple technique allows integrating in `ArcView` more complex processing. Then, using join indices could increase the extensibility of current data analysis system (as statistical tools).

6. Conclusion

In this paper, we emphasize that spatial relationships are of great importance in spatial data analysis and are used intensively in spatial data mining. This notion of spatial relationships does make the difference between data mining on alphanumerical data and on spatial data. It is represented by different concepts, such as contiguity matrix or neighborhood graphs, used in spatial analysis and spatial data mining.

In this paper, spatial relationships are materialized by an extension of the well-known join indices. In addition to its traditional use as an accelerator in the process of spatial join, this concept is deviated here in order to specify it as an efficient tool for spatial data mining. The definition of this tool is the main contribution of this paper. This tool is very simple to use thanks to its representation using the relational paradigm. Join indices can be handled in the same way than other tables and manipulated with the powerful and standardized SQL query language. We have also proposed a software architecture schema allowing the integration of such a structure into an existing system. Join indices could be handled as an external process of any system that do not provide any spatial functionality.

We have shown the great interest of the compatibility of this structure with contiguity matrices. Join indices can be viewed as an implementation of contiguity matrices allowing analysis on large sets of data. Once join indices are pre-computed, they may be used to apply analysis methods such as spatial auto-correlation, spatial characterization or spatial decision trees [10].

This paper also argues about the possibilities for coding join indices. A prototype have already be developed using the ArcView environment according to a SJI coding choice described in this paper. A second prototype is under construction with respect to the software architecture described in this paper. A short-term perspective is to compare in terms of performance the two different coding possibilities, i.e. OJI and SJI.

In general context of spatial data mining, there is many research issues. One issue concerns user assistance and user involvement in the process of knowledge discovery. Indeed, results interpretation could be sometimes a very hard task. Moreover, users should be able to intervene during the process to orient or to filter the analysis. The application of techniques coming from the field of visualization and interaction is already proposed in this objective [6, 17] in the context of conventional data mining. Extending these concepts to spatial data mining such as providing visualisation techniques for spatial relationships, including the user in the whole process of spatial data mining is an interesting perspective.

References

1. Andrienko, N. and Andrienko, G.: Interactive Maps for Visual Data Exploration, International Journal of Geographical Information Sciences 13 (4), pp. 355-374 (1999). See also URL: http://borneo.gmd.de/and/icavis.

2. Anselin, L.: Local indicators of spatial association - LISA. *Geographical Analysis*, 27, 2, pp. 93-115 (1995)
3. Bédard, Y., Lam, S., Proulx, M.J., Caron, P.Y. and Létourneau, F.: Data Warehousing for Spatial Data: Research Issues, Proceedings of the International Symposium Geomatics in the Era of Radarsat (GER'97), Ottawa (1997) pp. 25-30
4. Bennis K., David B., Quilio I., Thévenin J-M. and Viémont Y..: GéoGraph : A Topological Storage Model for Extensible GIS:, Proc. of Auto-Carto'10, Baltimore, USA, 368-392 (1991)
5. Burtschy, B. and Lebart, L.: Contiguity analysis and projection pursuit. In : *Applied Stochastic Models and Data Analysis*, R. Gutierrez and M.J.M. Valderrama, Eds, World Scientific, Singapore, pp. 117-128 (1991)
6. Card, S.K., Mackinlay, J.D. and Shneiderman, B.: Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann (1999)
7. Cliff A.D., Ord J.K.,: Spatial autocorrelation, Pion, London (1973)
8. Egenhofer M.J. and Sharma J.: Topological Relations Between Regions in R2 and Z2, Advance in Spatial Databases, 5th International Symposium SSD'93. pp 316-331. Singapore (1993) Springer-Verlag.
9. Ester, M., Frommelt, A., Kriegel, H.-P., Sander, J.: Algorithms for Characterization and Trend Detection in Spatial Databases", *Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining,* New York, NY (1998).
10. Ester, M., Kriegel, H.-P., Sander, J.: Spatial Data Mining: A Database Approach, Proceedings of the 5th Symposium on Spatial Databases, Berlin, Germany (1997)
11. Fayyad et al.: Advances in Knowledge Discovery and Data Mining, AAAI Press / MIT Press (1996)
12. Fisher, M. and Getis, A.: spatial analysis – spatial statistics, behavioural modelling and neurocomputing, Berlin, Springer (1997)
13. Fotheringham, S. and Rogerson, P.: Spatial Analysis and GIS, Taylor and Francis (1995)
14. Han J., Cai Y. and Cerone N.: Knowledge Discovery in Databases; An Attribute-Oriented Approach., Proceedings of the 18th VLDB Conference. Vancouver, B.C (1992) pp. 547-559. See also URL : http://www.cs.sfu.ca/~han
15. Han J., Koperski K., and Stefanovic N.: GeoMiner: A System Prototype for Spatial Data Mining, Proc. 1997 ACM-SIGMOD Int'l Conf. on Management of Data (SIGMOD'97), Tucson, Arizona, May1997 (System prototype demonstration).
16. Holt, D., Steel D.G., Tramer M.: Area Homogeneity and the Modifiable Areal Unit Problem, *Geographical Systems (3)*, pp. 181-200 (1996)
17. Keim, D.A., Kriegel, H.P.: Visualization Techniques for Mining Large Databases: A Comparison, IEEE Transactions on Knowledge and Data Engineering, vol 8, n°6 (1996)
18. Khoshafian, S.N, and Copeland, G.P: Object Identity. In Proc. of the ACM Conf. on Object-Oriented Programing Systems and Languages (OOPSLA), pages 408-416. (1986)
19. Kraak, M.J. and MacEachren, A.M.: Visualisation for exploration of spatial data. International Journal of Geographical Information Sciences 13 (4), pp. 285-287 (1999)
20. Kraak, M.J.: Visualizing spatial distributions. Chapter 11 in Longley, P., M. Goodchild, D. Maguire & D. Rhind (editors) Geographical information systems: principles, techniques, management and applications. New York: J. Wiley & Sons (1999) pp.157-173.
21. Laurini R., Thompson D.: Fundamentals of Spatial Information Systems, Academic Press, London, UK, 680 p, 3rd printing (1994)
22. Laurini, R.: Information Systems for Urban Planning : A Hypermedia Cooperative Approach , Taylor and Francis (2000)

23. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. : Geographic Information Systems , Volume 1, Wiley, 1999
24. Lu, W. and Han, J: Distance-Associated Join Indices for Spatial Range Search. Eighth International Conference on Data Engineering, (1992) Tempe, Arizona, pp. 284-292
25. Maier, D., The Theory of Relational Databases, Computer Science Press, 1983.
26. Matheron, G. : Principles of geostatistics. *Economic Geology*, 58, pp. 1246-1266, (1963)
27. O' Neil, P. and Graefe, G: Multi-tables joins through bitmapped join indices. SIGMOD Record, 24(3), pp. 8-11 (1995)
28. Openshaw, S., Charlton, M., Wyme,r C. and Craft, A: A mark 1 geographical analysismachine for the automated analysis of point data sets, International Journal of Geographical Information Systems, Vol. 1 (4), pp. 335-358 (1987). See also URL : http://www/ccg.leeds.ac.uk/smart/gam/gam.html
29. Roddick, J.F, Spiliopoulou, M.: A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research, ACM SIGKDD Explorations, volume 1, Issue 1 (1999)
30. Rotem D: Spatial join indices, Proc. of 7th Conf. on Data Engineering, Kobe, Japan (1991) pp. 500-509
31. Tobler W. R.: Cellular geography, In Gale S. Olsson G. (eds.) Philosophy in Geography, Dortrecht, Reidel (1979) 379-386
32. Valduriez P., "Join indices", *ACM Trans. on Database Systems*, 12(2); 218-246, June 1987.
33. Wang, W., Yang, J. and Muntz, R.: STING+: An approach to active spatial data mining, Proceedings of the Fifteenth International Conference on Data Engineering, Sydney, Australia. (1999) IEEE Computer Society. 116-12
34. Yeh, T-S: Spot: Distance based join indices for spatial data, ACM GIS 99, Kansas City, USA, pp 103-110 (1999)
35. Zeitouni K.: A Survey on Spatial Data Mining Methods Databases and Statistics Point of Views, Information Resources Management Association International Conference (IRMA'2000), Data Warehousing and Mining Track, Anchorage, Alaska, USA (2000)