UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

DIPARTIMENTO DI INFORMATICA

Relational Learning from Spatial Data: Retrospect and Prospect

Donato Malerba

malerba@di.uniba.it

EGC 2012 – 12e Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances

Bordeaux, France, 31 janvier - 3 février, Bordeaux, France

Outline 1/2

- Pervasiveness of spatial data
- Spatial vs. Traditional Data Mining
  - Heterogeneity + implicit spatial relationships defined by locational properties + spatial correlation;
- Spatial models
  - spatial lag, spatial error, spatial cross-regressive
- Opportunities for a Relational Approach

Outline 2/2

- Retrospect and prospect on some challenges:
  - Dynamic handling of spatial dependencies & scalability
  - Dealing with correlation and autocorrelation
  - Collective inference
  - Exploiting unlabelled data
  - Hierarchies of objects
  - Mining changes in space and time
- Collaboration is needed!

Pervasiveness of Spatial Data

- *1990 report from the Ohio Geographically Referenced Information Program (OGRIP):* 80% of all data has a *spatial* dimension
- GPS, high-resolution remote sensing, location-aware services

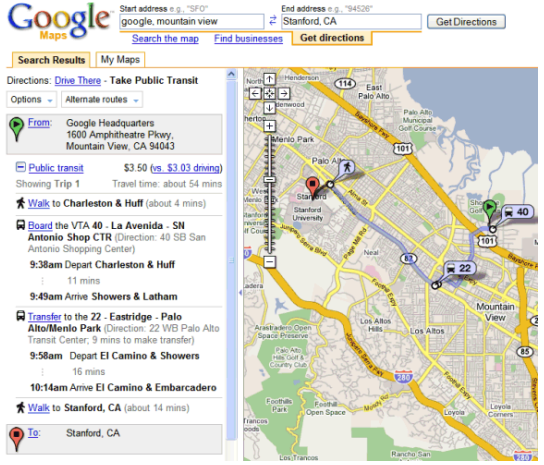
+

- Mapping websites, such as Google Maps, Live Search Maps and Yahoo Maps
- Today 95% is a more accurate estimate

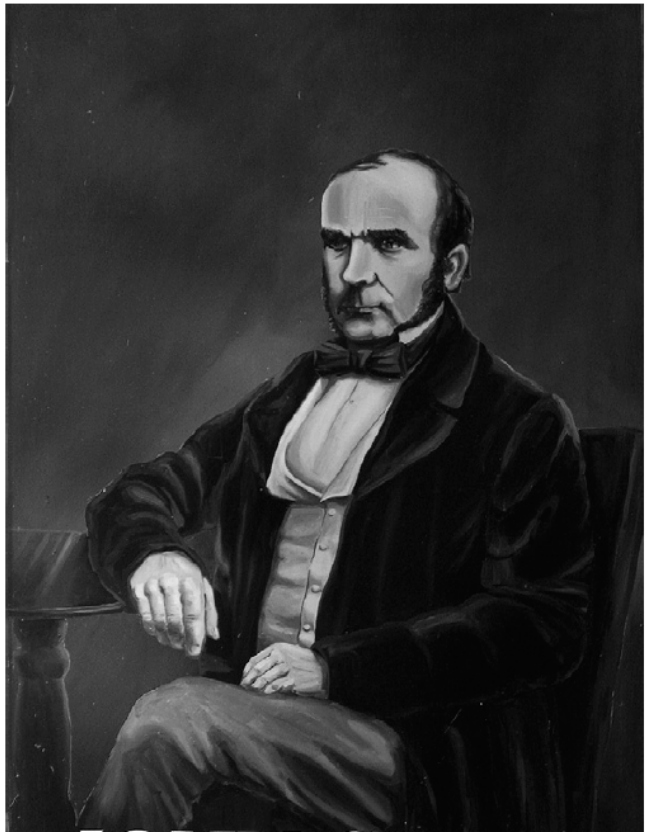
(*Stuart Hamilton*, GIS program director at the College of William and Mary, Williamsburg, VA)

Mapping Data

- By connecting public or company data to a map, it is possible to discover new patterns.
- Map-based presentations of patterns are more effective than charts and graphs.
- Data from most sciences can be analyzed “spatially”.



Mapping Data: A Historical Example

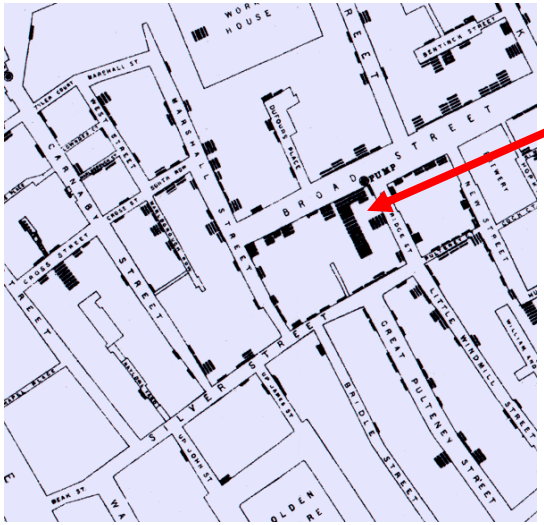


John Snow

- 1848: An epidemic of the ‘Asiatic cholera’ hit London
- John Snow observed the distribution of deaths throughout the city and
- hypothesized that river water contaminated by cholera evacuations explained spatial variations in mortality throughout London

Mapping Data: A Historical Example

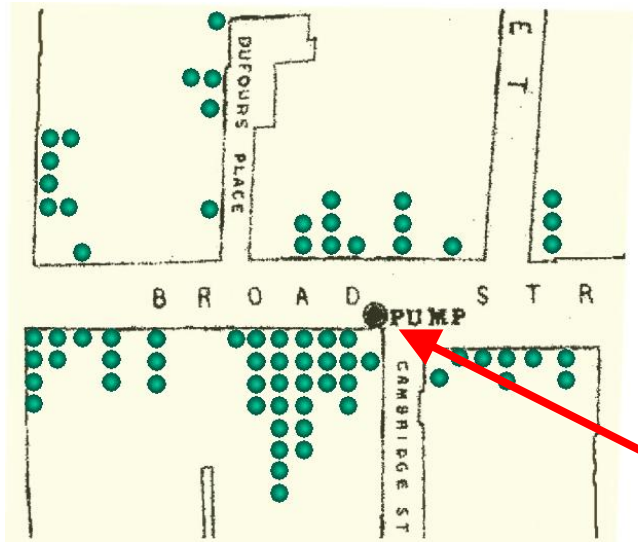
- August 1854: the cholera epidemic hit an area of North London
- J. Snow obtained the names and addresses listed on 83 death certificates from the Registry Office.



He marked cholera cases on a map

Mapping Data: A Historical Example

- He also inventoried potential sources of contamination (pumps)
- and combined this information on the map.



Nearly all the deaths within a short distance of the pump in Broad Street

Mapping Data: A Historical Example

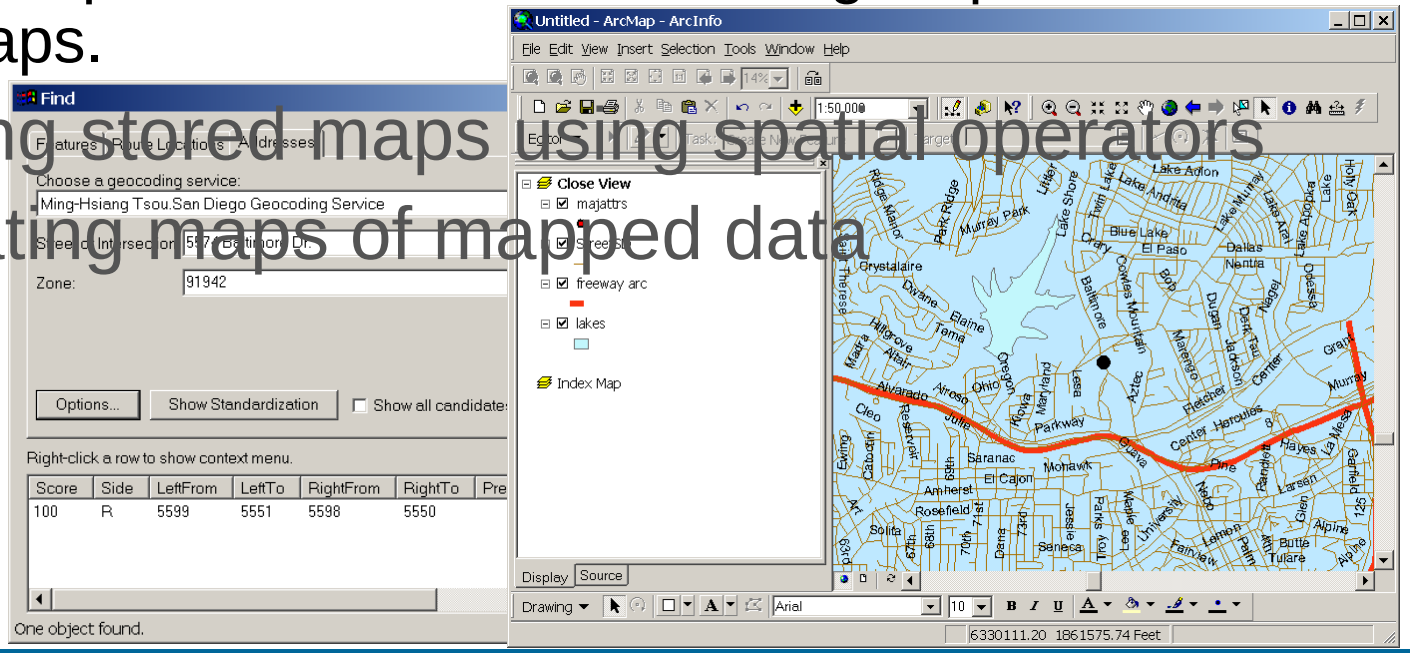
- Snow persuaded the parish council to remove the handle.
- Difficult decision: water provided by this pump was held in such high esteem that people came from neighboring streets for it.
- Result: the epidemic subsided.

Mapping Data: A Historical Example

- A curiosity: Snow's theory was supported by two pieces of '**negative data**'
  - No infection in the workhouse (it had its own well)
  - No cases in the Lion Brewery (workers drank beer)

Geographical Information Systems

- Popular technology that allows users to
  - represent thematic maps (in spatial databases)
  - Map non-spatial data
    - Convert patients' addresses to the geospatial location on maps.
  - Querying stored maps using spatial operators
  - Generating maps of mapped data

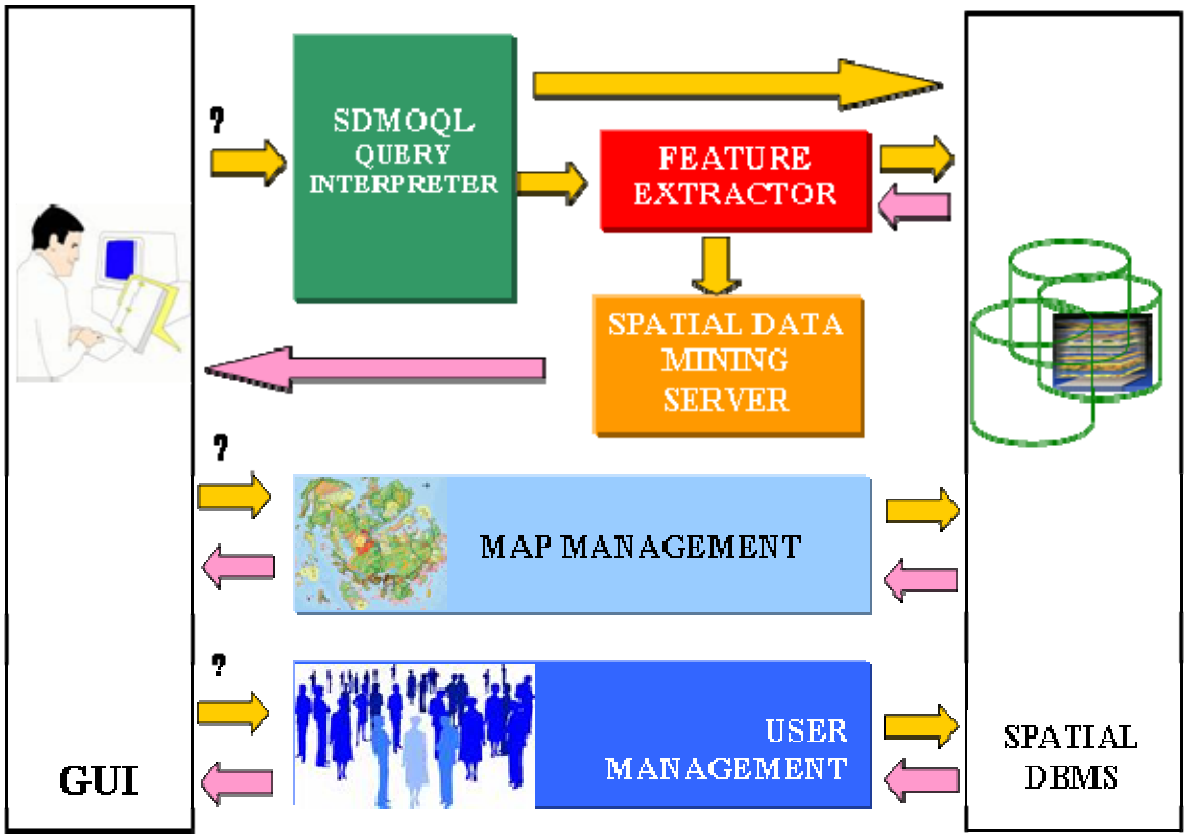


Geographical Information Systems

- Spatial visualization of individual variables is effective but overlooks complex multivariate dependencies.

Integrate GIS with
spatial data mining tools!

INGENS



D. Malerba, A. Lanza, & A. Appice (2009). Leveraging the power of spatial data mining to enhance the applicability of GIS technology.

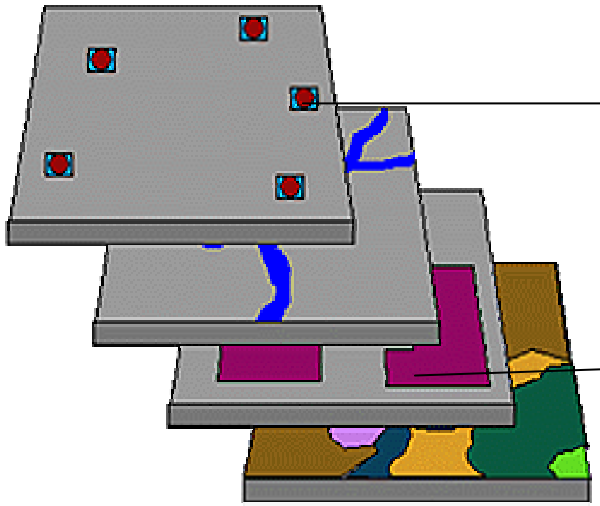
Spatial vs. Traditional Data Mining

- Features of spatial information that make it special:

1. Heterogeneity of spatial objects;

Heterogeneity

- Different types of spatial objects ➔ Several layers in a spatial DB



wells

ID	Location	Depth	Type
2357	(15,22)	20	drilled

buildings

ID	Surface	Size	Owner
AD18		250	Smith

Spatial vs. Traditional Data Mining

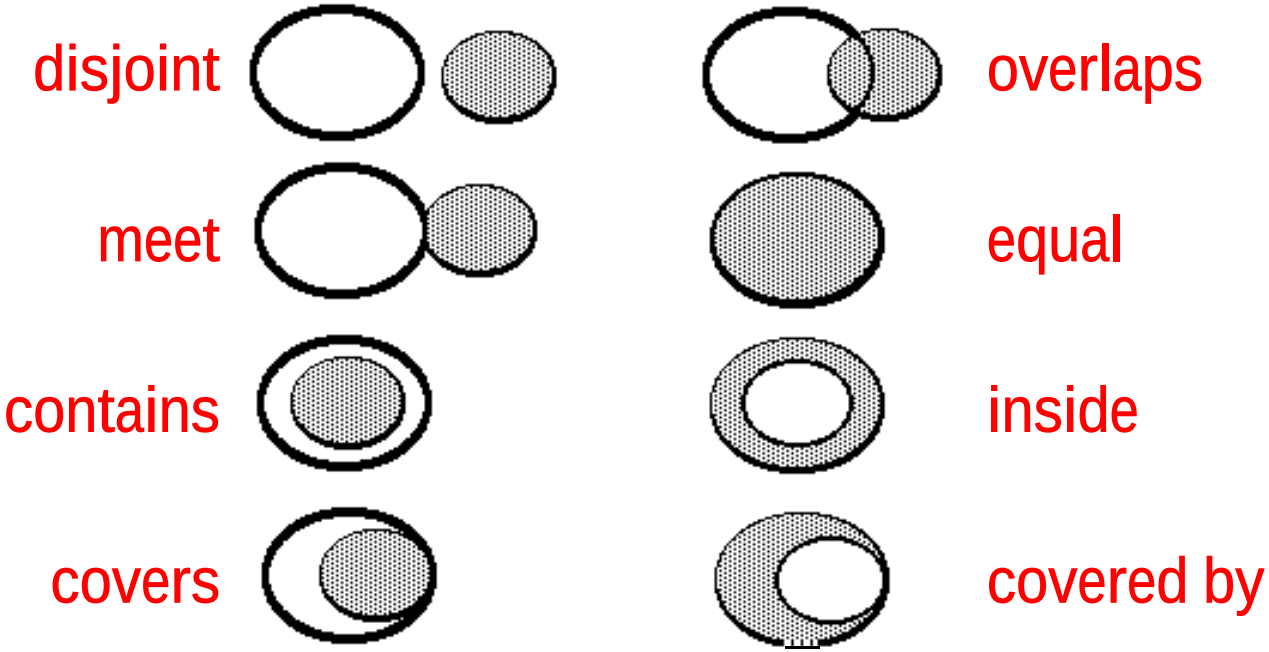
- Features of spatial information that make it special:

1. Heterogeneity of spatial objects;
2. Spatial objects have a locational property which implicitly defines several spatial relationships;

Topological Relationships

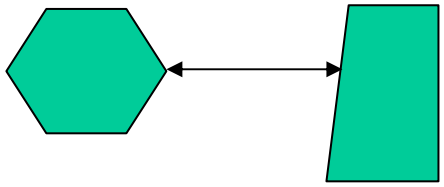
- Invariant under homomorphisms (rotation, translation & scaling)
- Semantics defined by the 9-intersection model

For regions



Distance Relationships

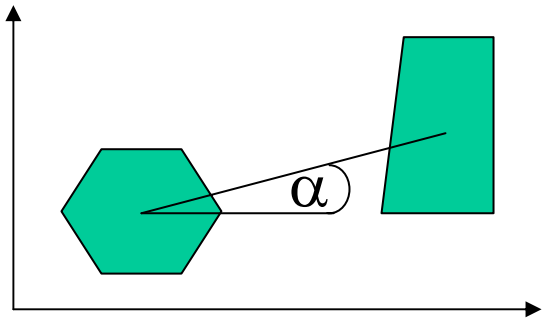
- Metric
  - Euclidean distance between two points
  - For polygons it's an aggregate function (e.g., minimum)



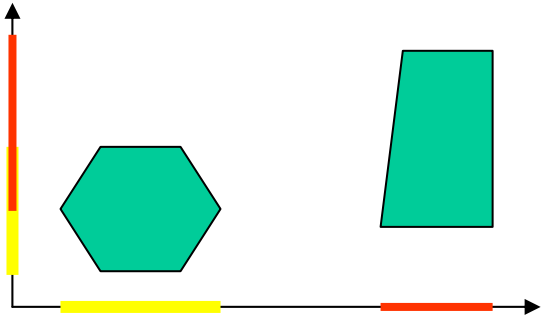
- Non-metric
  - Typically defined on the basis of a cost function (e.g. drive time)

Directional Relationships

- Based on an angle

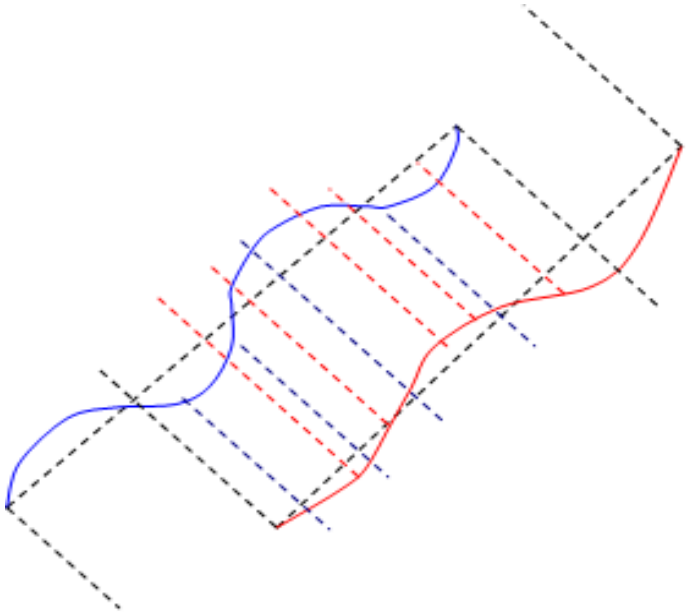


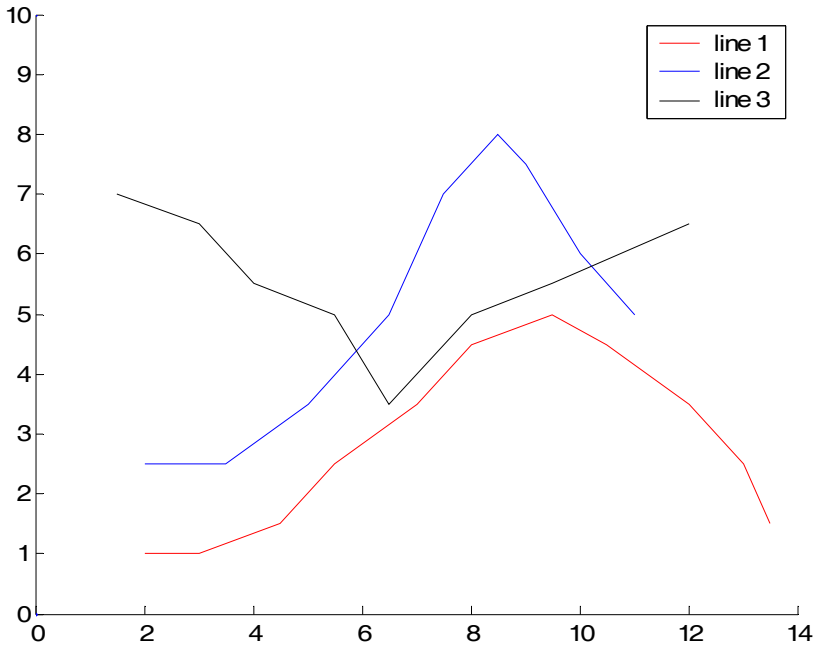
- Based on the extension of Allen's algebra



Hybrid Relationships

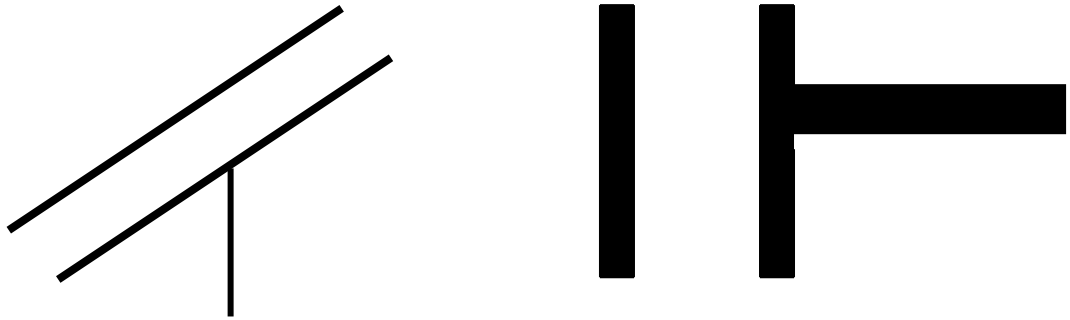
- Line parallelism (combination of a topological + distance relationship)





Abstraction from Physical Representation

- Interest towards properties not related to physical representation
  - Example: Two roads can cross each other, or run parallel, or can be confluent, independently of the fact that they are represented as “lines” or “regions”



Spatial vs. Traditional Data Mining

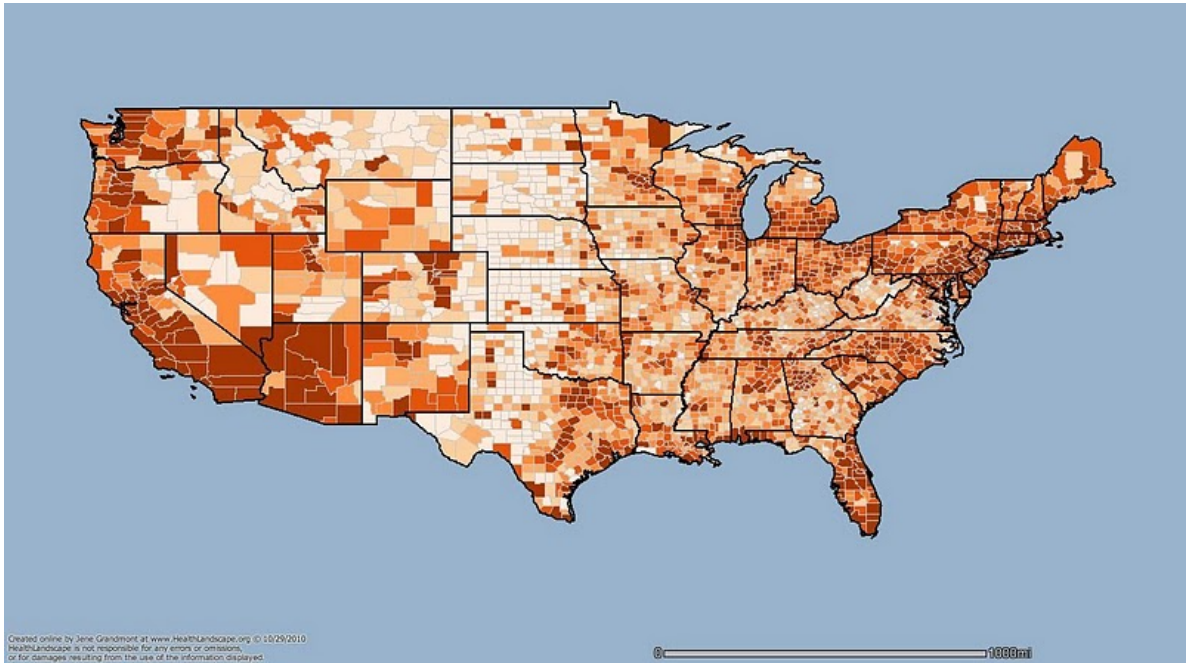
- Features of spatial information that make it special:

1. Heterogeneity of spatial objects;
2. Spatial objects have a locational property which implicitly defines several spatial relationships;
3. Attributes of spatially interacting objects tend to be statistically correlated.

Spatial Cross-correlation

- Correlation between two distinct attributes across space

Number of Children Ages 5 to 13 (Population Estimates), 2009)

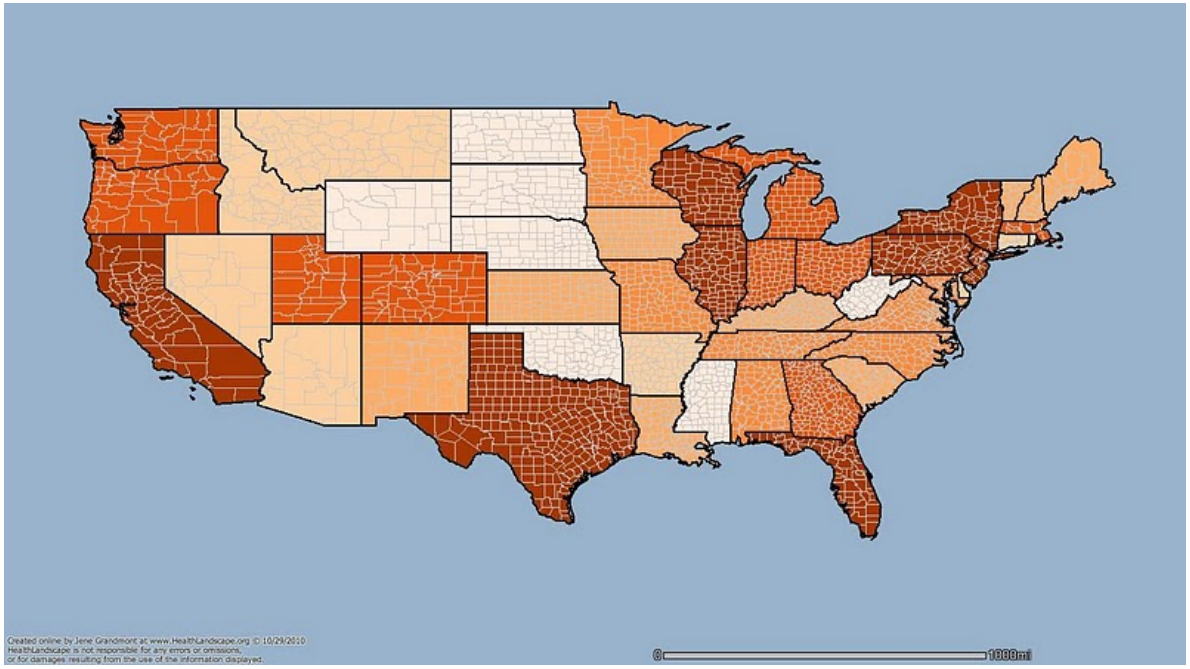


Courtesy of “The Health Foundation”

Spatial Cross-correlation

- Correlation between two distinct attributes across space

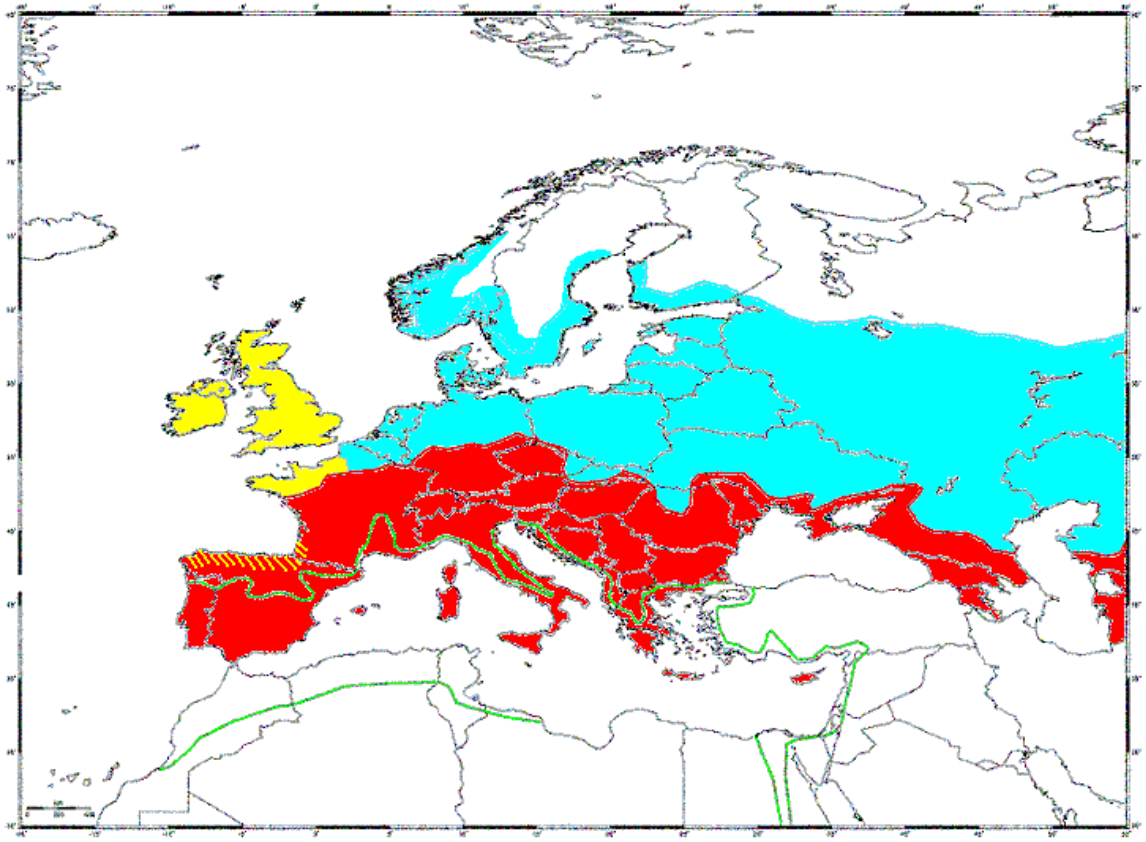
Number of Confectionery Manufacturing Establishments (Chocolate and Non-Chocolate), 2008



Courtesy of “The Health Foundation”

Auto-correlation

- Correlation of an attribute with itself across space



Auto-correlation: Justifications

- Tobler's first law of geography: Everything is related to everything else, but nearby things are more related than distant things.
- Homophily in social networks: a node of a specific class is likely to link to another node of the same class.

Auto-correlation: Positive vs. Negative

- Positive: in v, the attribute Y takes a value similar to those in N(v).
- Negative: in v, the attribute Y takes a value different from those in N(v).

Positive autocorrelation has received most of the attention, largely because few empirical examples of global negative autocorrelation have been found in spatial and social phenomena.

Spatial Relations vs. Statistical Correlation

- The network of data defined by implicit spatial relationship ➔ the network of spatial dependencies

  distance of houses from main roads is a necessary condition for autocorrelation of burglaries

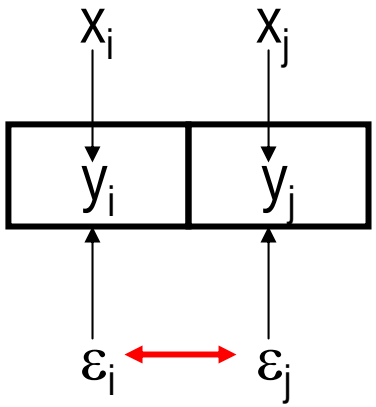
- Spatial correlation ➔ selection of relevant spatial relationships

  cross-correlation between price level of houses and the quality of services available in the nearby shows that the distance between houses and services is an important spatial relationship

Spatial Dependence

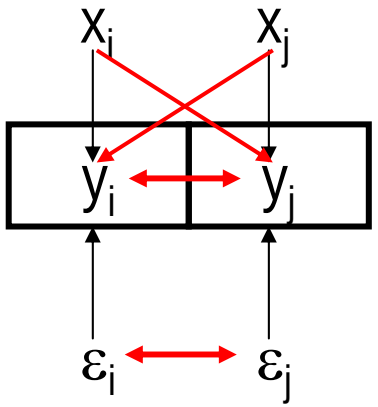
- Error
- Explanatory (non-target) variables
- Response (target) variables

Spatial error lag



Spatial

Spatially lagged explanatory variables



Spatially lagged response variables

Violated Assumptions

- Spatial error: error terms are uncorrelated
- Spatial lag: observations are independent (as well as error terms are uncorrelated)

"Anyone seriously interested in prediction when the sample data exhibit spatial dependence should consider a spatial model"

(LeSage & Page, 2001)

Models Developed in Statistics

1. spatial lag model: autocorrelation
2. spatial error model: correlation of errors
3. spatial cross-regressive model: cross-correlation.

L. Anselin, A., Bera, A. (1998): Spatial dependence in linear regression models with an application to spatial econometrics.

General Spatial Model

$$y = X\alpha + \beta D_1 y + \gamma D_2 X + u$$

- y: vector of observations of the dependent variable
- X: matrix of observations of the independent variables
- α: strength of local influence
- β: strength of autocorrelation
- γ: strength of cross-correlation
- D_1, D_2: spatial weight matrices (or *neighborhood matrices*)

General Spatial Model

$$u = \lambda D_3 u + \varepsilon$$

- u: error
- λ: strength of error (auto-)correlation
- D_3: spatial weight matrix
- $\varepsilon \sim N(0,\sigma^2)$; (*homoskedasticity*)

General Spatial Model

$\beta = 0$, $\gamma = 0$, $\lambda = 0$: **classical linear model**

What are the determinants of the price of a house?

Price = Sq. mt. + Age + Median Income + Dist. to Metro + error

General Spatial Model

$\gamma = 0$, $\lambda = 0$: **spatial lag model**

How do property appraisers determine the value of a property?

Price = D_1 * Price + Sq. mt. + Age + Median Income + Dist. to Metro + error

General Spatial Model

$\beta = 0$, $\gamma = 0$: **spatial error model**

What are the determinants of the price of a house?

Price = Sq. mt. + Age + Median Income + Dist. to Metro + error

error = D_3*error + ε

Spatial error can occur when:

1) Spatially correlated omitted variables
2) Spatially correlated aggregate variables
3) Spatially correlated errors in variable measurement

General Spatial Model

$\beta = 0$, $\lambda = 0$: **spatial cross-regressive model**

What are the determinants of the price of a house?

Price = Sq. mt. + Age + D_2*Age + Median Income + Dist. to Metro + error

Spatial Modeling: Limitations

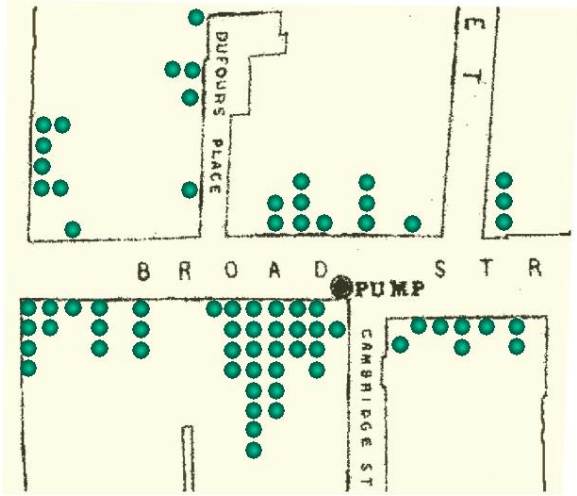
- D has to be carefully defined
- How can D express the contribution of different spatial relationships?
- Spatial dependencies are all handled in a pre-processing or feature extraction step
- All spatial objects involved in the spatial phenomena (rows of X) are uniformly represented by the same set of attributes
- No clear difference between …

Reference vs. Task-relevant Objects

- **Reference objects**: the main subject of analysis
- **Task-relevant objects**: objects in the neighborhood that can contribute to explain the spatial variation

Example:

- **Reference objects**: buildings where cholera cases are registered
- **Task-relevant objects**:
  - pumps
  - shops
  - …

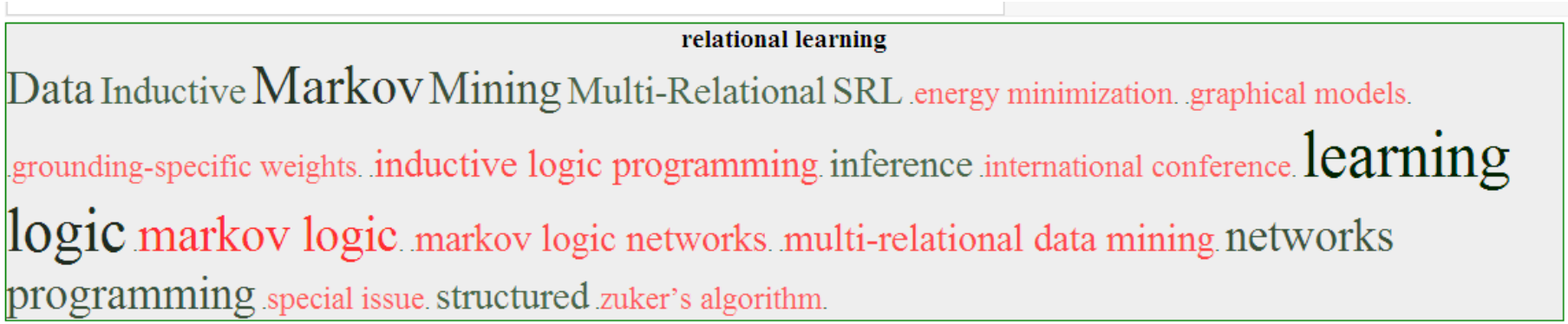


Recap

- Spatial data are ubiquitous
- Visualization tools provided by GIS technology uncover simple spatial dependencies
- Mining spatial data presents some issues
  - Heterogeneity
  - Many implicitly defined spatial relationships
  - Spatial correlation (cross- or auto-)
- Spatial models present limitations (and are specific of predictive modeling)

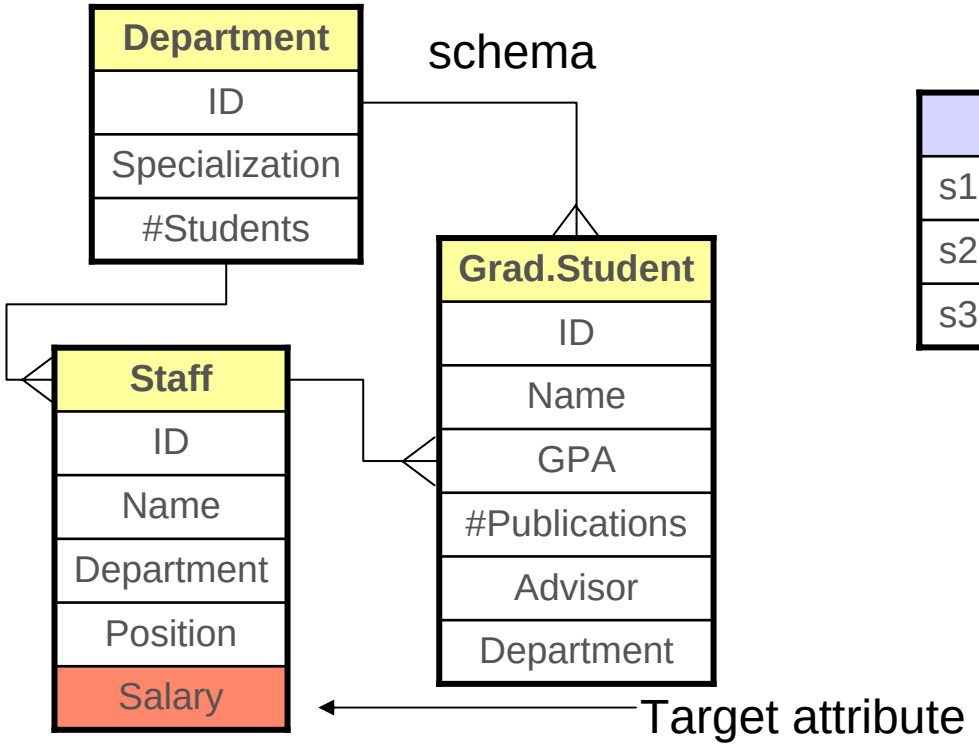
Opportunities for a Relational Approach

- Relational mining algorithms can be directly applied to various representations of networked data, i.e. collections of interconnected entities.
- Investigated under many names



Relational approach: a DB view

Units of analysis are represented a network of relations (links are foreign key constraints).



Department		
d1	Math	1000
d2	Physics	300
d3	Computer Science	400

Graduate Student					
s1	John	2.0	4	p1	d3
s2	Lisa	3.5	10	p4	d3
s3	Michel	3.9	3	p4	d4

observations

Staff				
p1	Dale	d1	Professor	70 - 80k
p2	Martin	d3	Postdoc	30-40k
p3	Victor	d2	Visitor Scientist	40-50k
p4	David	d3	Professor	80-100k

(M)RDB and FOL terms

There is a strict correspondence between (M)RDB and First-order Logic (FOL)

- Relation name R
- Attribute of relation
- Tuple $(a_1, \ldots, a_n)$ or relation R
- Relation R as set of tuples
- Relation R defined as a view

- Predicate symbol r
- Argument of predicate
- Ground fact r $(a_1, \ldots, a_n)$
- Predicate r defined extensionally
- Predicate r define intensionally

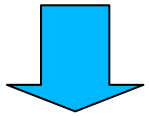
Opportunities for a Relational Approach

➔ Relational mining can consider various forms of correlation which bias learning in spatial domains.

➔Relational patterns reveal spatial dependencies and which of the many spatial relations are relevant.

Opportunities for a Relational Approach

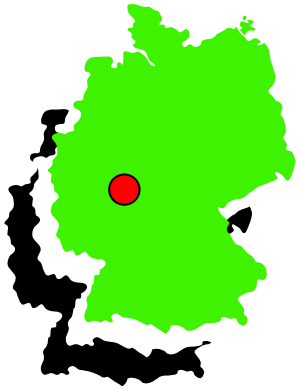
- Spatial database as a kind of networked data where:
  - entities are spatial objects and
  - connections are spatial relationships



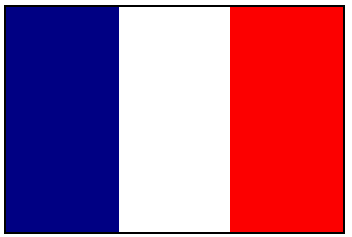
the application of relational mining methods is straightforward (at least in principle)

Background work: (some) geo-References

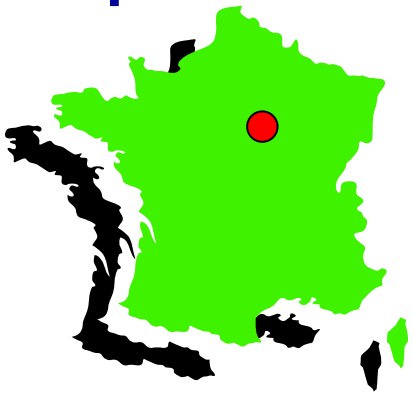
- **W. Klösgen, M., May (2002). Spatial subgroup mining integrated in an object-relational spatial database.PKDD**



Background work: (some) geo-References

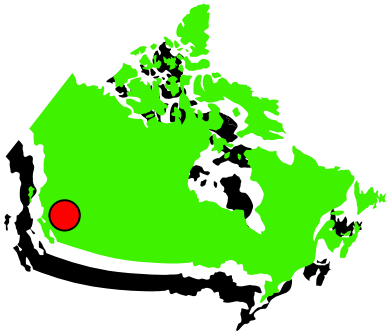


- **N. Chelghoum, K. Zeitouni (2004). Spatial data mining implementation: Alternatives and Performances. Spatial Data Mining**
- **N. Chelghoum, K. Zeitouni (2004). Mise en oeuvre des méthodes de fouille de données spatiales - Alternatives et performances. EGC**

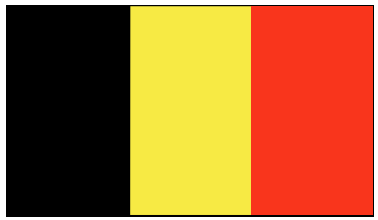


Background work: (some) geo-References

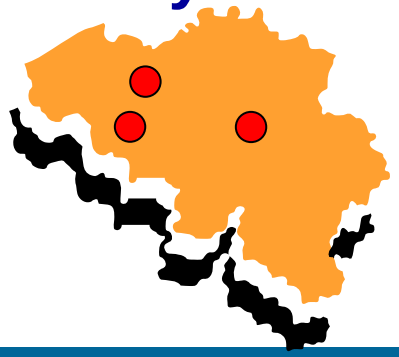
- **R. Frank, M. Ester, A.J. Knobbe (2009). A multi-relational approach to spatial classification. KDD.**



Background work: (some) geo-References



- **J. Maervoet, P. De Causmaecker, A. Nowé, G. Vanden Berghe (2008): Feasibility study of applying descriptive ILP to large geographic databases. In Workshop on mining multidimensional data.**
- **J. Maervoet, C. Vens, G. Vanden Berghe, H.Blockeel, P. De Causmaecker (2012). Outlier detection in relational data: A case study in geographical information systems . Expert Systems with Applications**

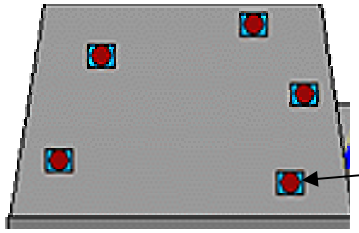


Background work: (some) geo-References

- A. Appice, M. Ceci, A. Lanza, F.A. Lisi, D. Malerba (2003): **Discovery of spatial association rules in geo-referenced census data: A relational mining approach**. Intell. Data Anal.
- M. Ceci, A. Appice, D. Malerba (2004): **Spatial Associative Classification at Different Levels of Granularity: A Probabilistic Approach**. PKDD
- D. Malerba, A. Appice, A. Varlaro, A. Lanza (2005): **Spatial Clustering of Structured Objects**. ILP
- D. Malerba, M. Ceci, A. Appice (2005): **Mining Model Trees from Spatial Data**. PKDD
- M. Ceci, A.Appice, D. Malerba (2007): **Discovering Emerging Patterns in Spatial Databases: A Multi-relational Approach.** PKDD
- D. Malerba (2008). **A relational perspective on spatial data mining**. IJDMMM 1(1), 103–118
- D. Malerba, M. Ceci, A. Appice: **A relational approach to probabilistic classification in a transductive setting**. Eng. Appl. of AI 22(1): 109-116

Integration with Spatial Databases

- Spatial data are generally stored in spatial databases (e.g. Oracle Spatial, Spatial Posgres), where a spatial object is represented by one or more tuples of a …
- **Layer**: a database relation R_i with a number of elementary attributes A^i_1, …, A^i_{mi} and possibly a geometry attribute G_i represented in vector mode.



wells

ID	Location	Depth	Type
2357	(15,22)	20	drilled

Integration with Spatial Databases

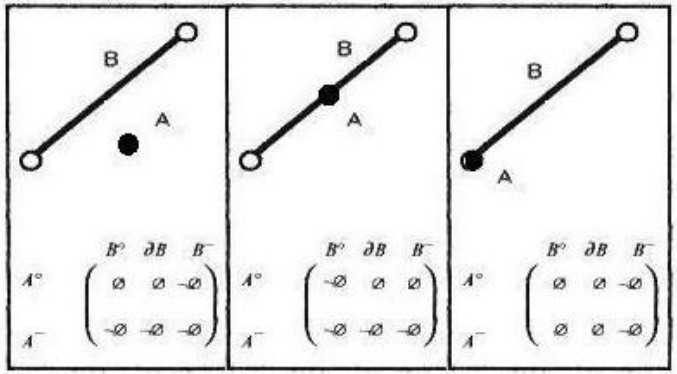
- Spatial relationships are implicit and, to be exhibited, they require costly joins on spatial criteria

$$R_i \bowtie_\rho R_j$$

  spatial relation ρ
- Too many spatial joins implicitly defined

Integration with Spatial Databases

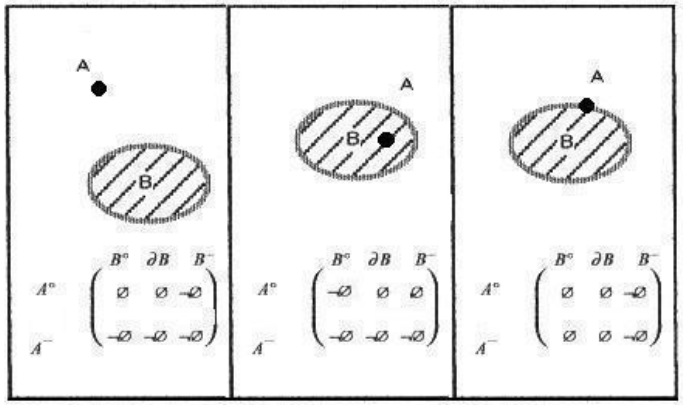
- Just an example: topological relations



Between a point and a line

Integration with Spatial Databases

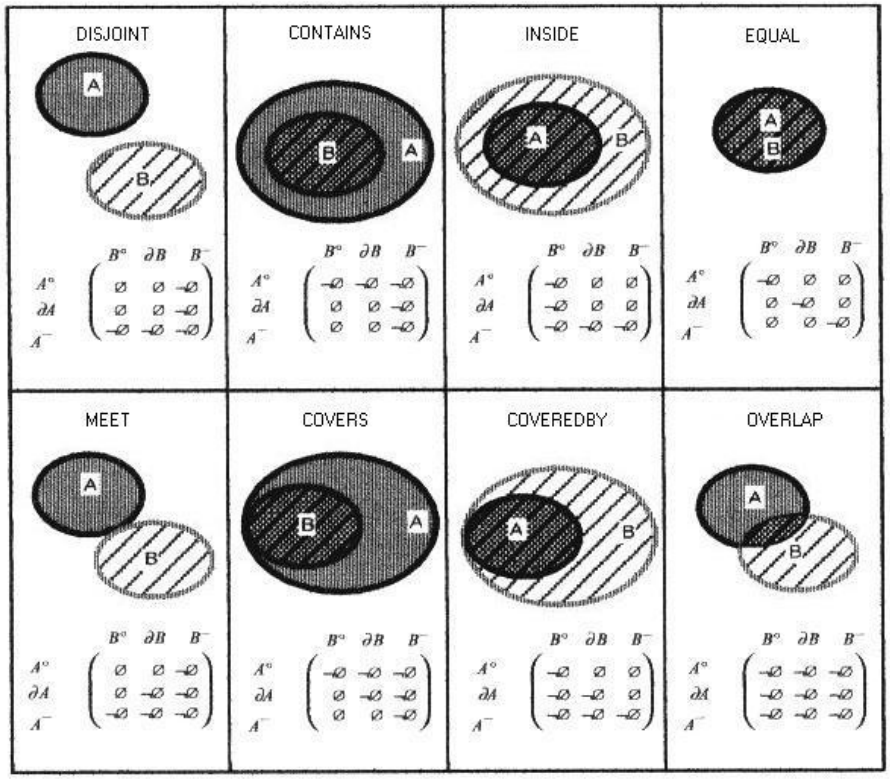
- Just an example: topological relations



Between a point and a region

Integration with Spatial Databases

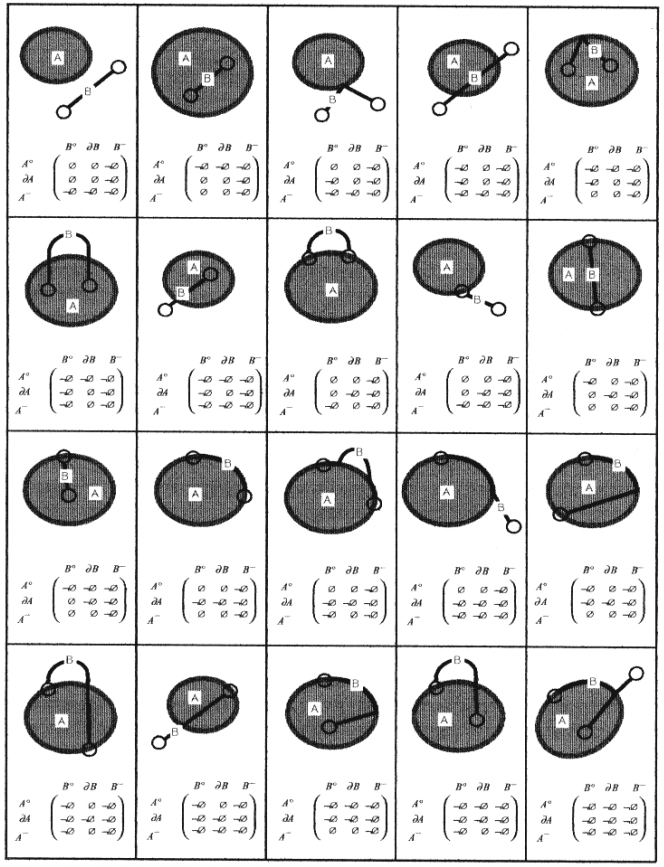
- Just an example: topological relations



Between two regions

Integration with Spatial Databases

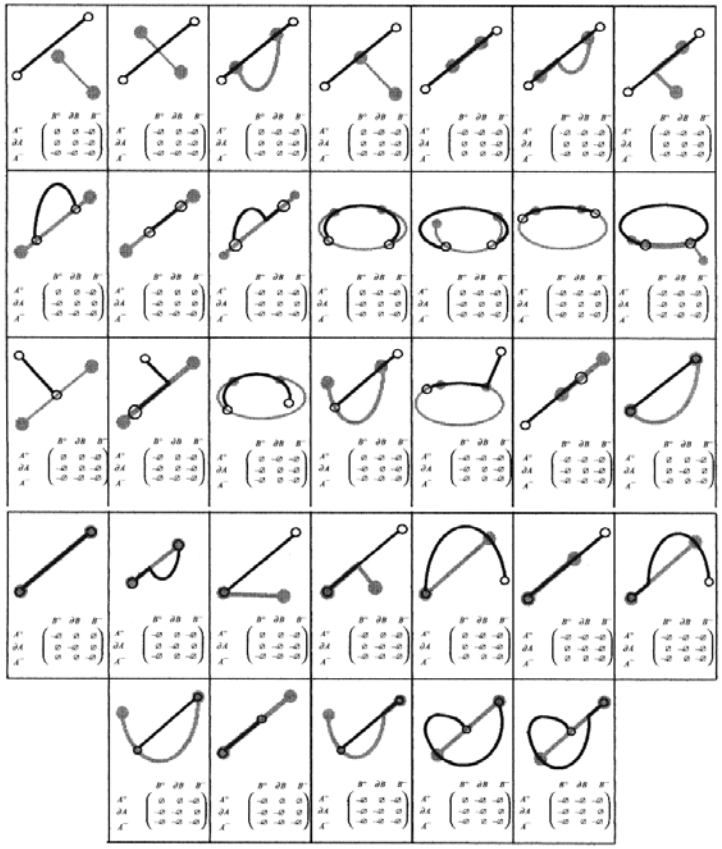
- Just an example: topological relations



Between a region and a line

Integration with Spatial Databases

- Just an example: topological relations



Between two lines

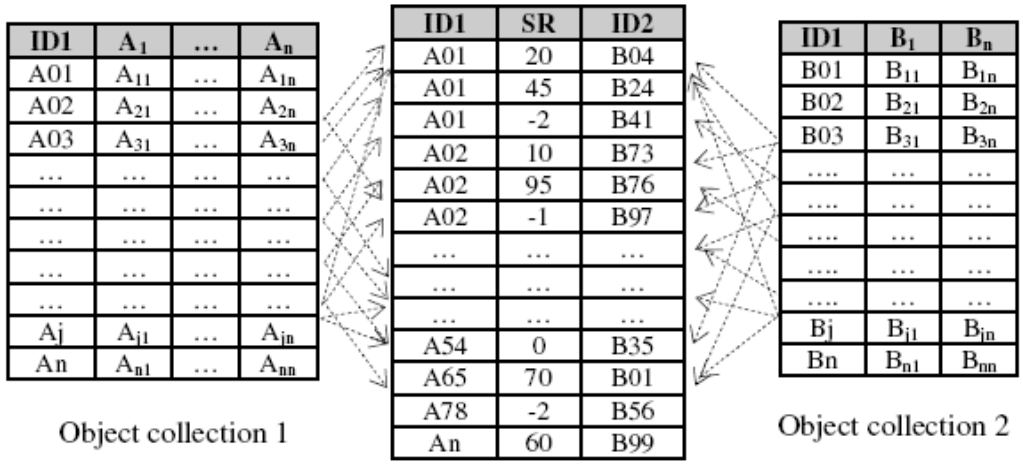
Integration with Spatial Databases

- Without the support of spatial database technology no spatial data mining is possible.
- SDBMS supports the creation of spatial indexes
  - Quadtrees
  - Kd-tree

  which are used to make the computation of spatial joins faster.
- Not enough!

Integration with Spatial Databases

- Chelghoum & Zeitouni study:



ID1	A_1	…	A_n
A01	A_{11}	…	A_{1n}
A02	A_{21}	…	A_{2n}
A03	A_{31}	…	A_{3n}
…	…	…	…
…	…	…	…
…	…	…	…
…	…	…	…
…	…	…	…
Aj	A_{j1}	…	A_{jn}
An	A_{n1}	…	A_{nn}

Object collection 1

ID1	SR	ID2
A01	20	B04
A01	45	B24
A01	-2	B41
A02	10	B73
A02	95	B76
A02	-1	B97
…	…	…
…	…	…
…	…	…
A54	0	B35
A65	70	B01
A78	-2	B56
An	60	B99

Spatial join index

ID1	B_1	B_n
B01	B_{11}	B_{1n}
B02	B_{21}	B_{2n}
B03	B_{31}	B_{3n}
….	…	…
….	…	…
….	…	…
….	…	…
….	…	…
Bj	B_{j1}	B_{jn}
Bn	B_{n1}	B_{nn}

Object collection 2

- N. Chelghoum, K. Zeitouni (2004). Spatial Data Mining Implementation: Alternatives and performances.

Integration with Spatial Databases

- First alternative: Querying on the fly
- Three tables in input:
  - target objects,
  - neighbor objects,
  - spatial join index.
- Whenever, the attribute to analyze is a neighborhood attribute, the algorithm does two join operations between the three tables.

Too expensive!!

Integration with Spatial Databases

- Second alternative: Join materialization
- Materializes, once for all, the joins on keys between the three tables.
- This avoids the multiplication of the joins queries but leads to the **duplication** of the analyzed objects.
- The mining algorithm should be modified (e.g., computation of information gain) to consider this.

Not easy!!

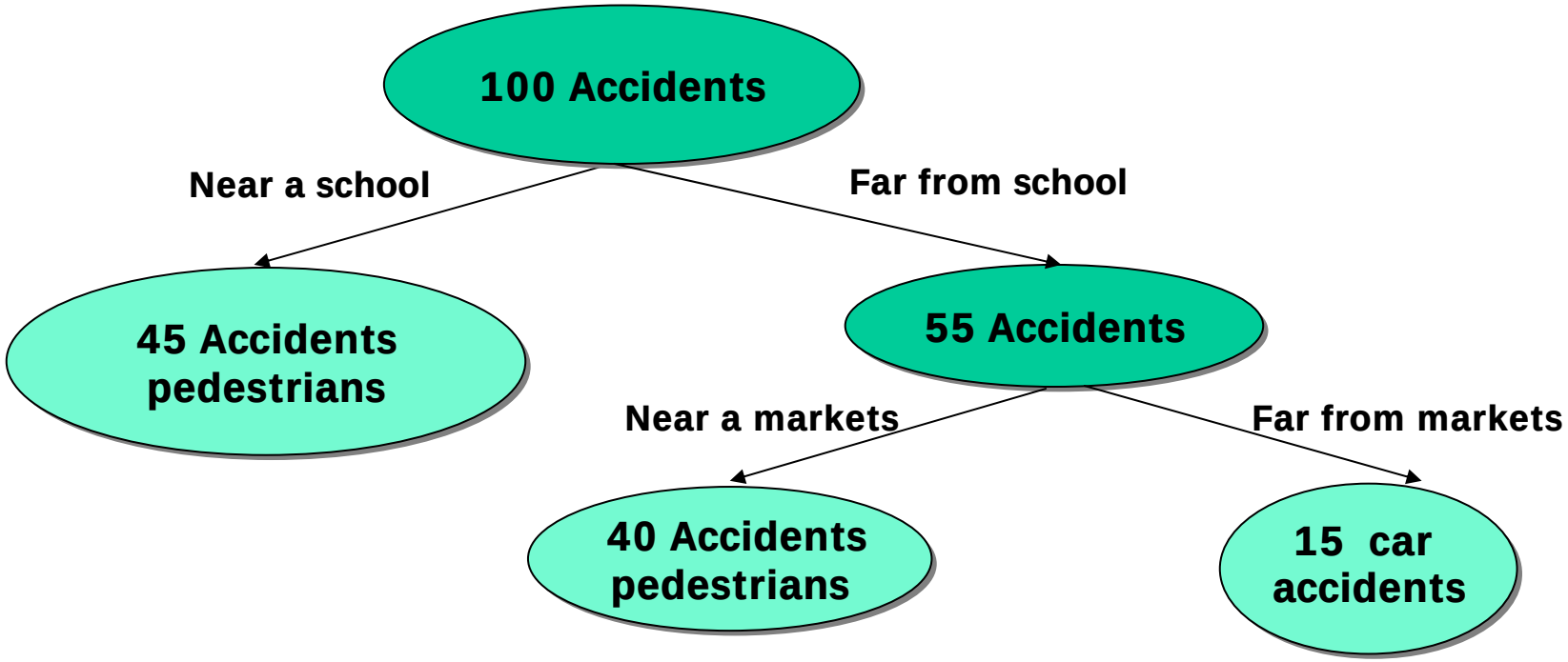
Integration with Spatial Databases

- Third alternative: Reorganize the data
- **Complete** the target table by the data present in others tables.
- **Aggregations** are used in case of multiple links

How to aggregate?

Integration with Spatial Databases

- Results for Spatial Decision Trees



Integration with Spatial Databases

- Results for Spatial Decision Trees

			1st alternative	2nd alternative			3rd alternative		
A	B	C	D	E	F	G	H	I	J
122	147	37	240	3	1	2	3	1	2
204	148	52	300	4	1	3	3	1	2
1594	6330	869	16860	7	1	6	9	5	4
3437	20180	869	24799	8	1	7	76	71	5
8668	20180	869	35205	18	2	16	157	150	7
15574	27054	869	48403	19	2	17	640	626	14
21892	53631	869	70020	56	2	54	925	882	43
29810	74302	869	88320	32	2	30	1372	1330	42

A: Size of R (objects)
B: Size of I (objects)
C: Size of V (objects)
D: Total time of the 1st alternative (seconds)
E: Total time of the 2nd alternative (seconds)
F: execution time of the first step (seconds)
G: execution time of the second step (seconds)
H: Total time of the 3rd alternative (seconds)
I: execution time of the first step (seconds)
J: execution time of the second step (seconds)

Integration with Spatial Databases

Research needs …

- No systematic study on various alternatives
- Relational mining methods take advantage of knowledge on the data model (e.g., foreign keys) to guide the search process, but …

  in spatial databases the navigation is based on spatial relationships which are not explicitly modeled.

Tight vs. Loose Integration

- Guarantees the applicability to large datasets
- Exploits knowledge of the available data model
- Avoids useless preprocessing of spatial relationships when there is no statistical correlation

SubgroupMiner (Kloesgen & May)

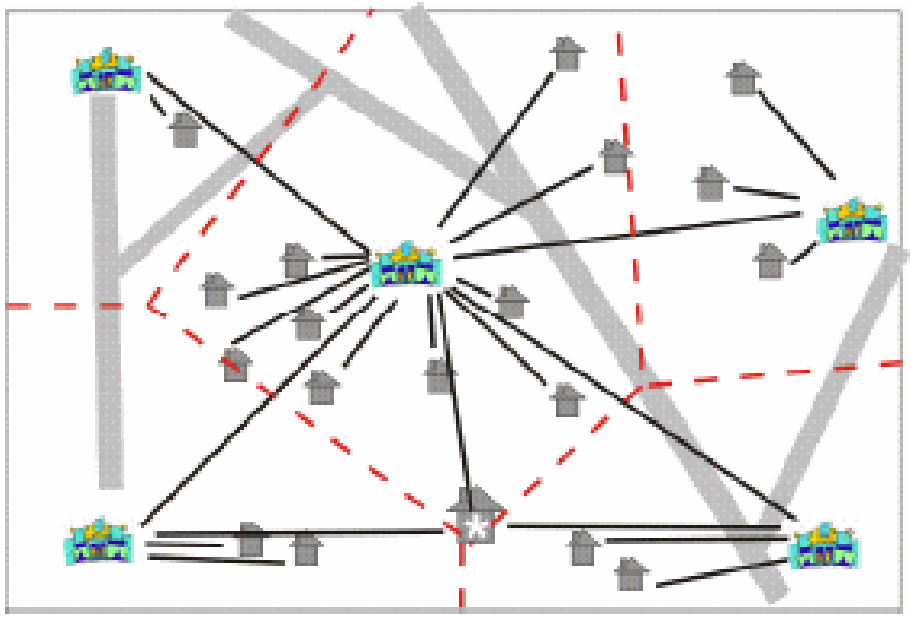
Mrs-SMOTI (Malerba et al)

- Focus on general aspects of the mining task
- Exploit important theoretical and empirical results

ARES (Appice et al)

Filtering out Irrelevant Relationships

- Frank, Ester, Knobbe (2009).
- Use of Voronoi diagrams to partition the space into cells and base on this the neighborhood relation between spatial objects.



Filtering out Irrelevant Relationships

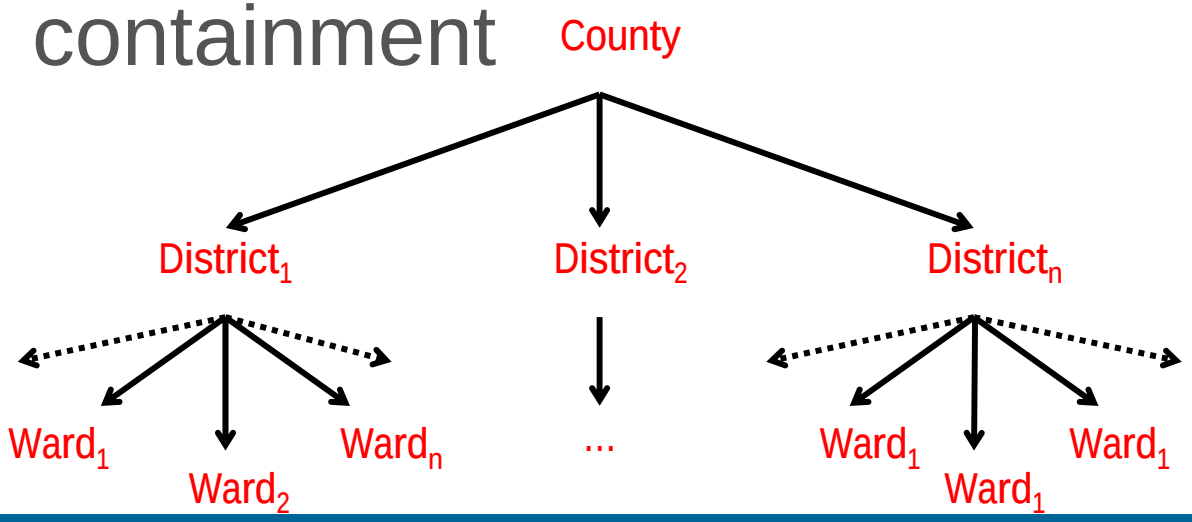
- Only one spatial relationship (neighbor).
- Why not considering the other spatial relationships? (e.g. spatial containment)

Research needs …

- Domain knowledge helps but can we use some coarse computation of spatial dependencies (cross or autocorrelations) to filter our irrelevant spatial relationships?

Hierarchical Representations

- Both spatial object and spatial relationships are organized in taxonomies which are typically represented by hierarchies.
- A hierarchy of areal objects may also be induced by the spatial relationship of containment



A spatial hierarchy for UK census data

Hierarchical Representations

- Spatial patterns involving the most abstract spatial objects are
  - easier to discover (due to coarser representations)
  - well supported, but
  - less confident

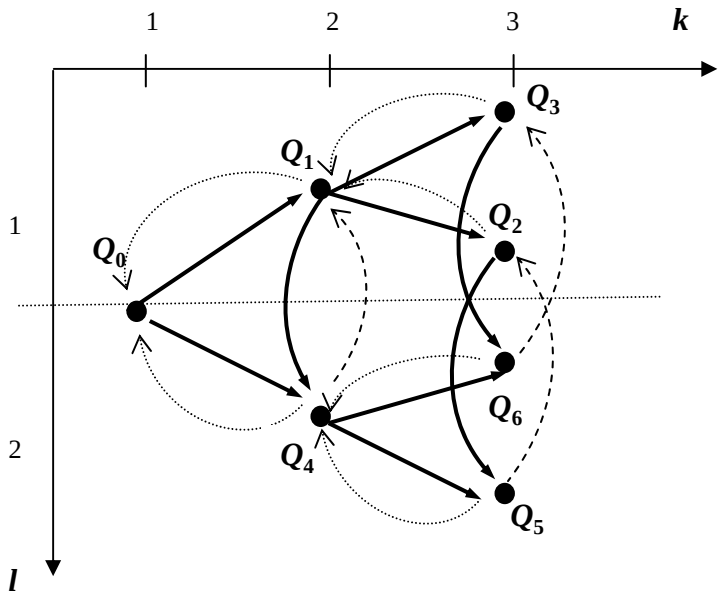
- Spatial data mining methods should be able to explore the search space at different granularity levels.

Hierarchical Representations

- Naive approach:
  - Level-by-level analysis
  - Information on patterns found at a level is not used to make search more efficient at a higher/lower level
- More sophisticated approach:
  - GeoAssociator (Koperski & Han, 1995)
  - SPADA (Malerba & Lisi, MLJ 2004)

Hierarchical Representations

SPADA (Malerba & Lisi, MLJ 2004)



Q_0 = q(X) ← & X:LargeTown

Q_1 = q(X) ← intersects(X,Y) &
X:LargeTown, Y: Road

Q_2 = q(X) ← intersects(X,Y), intersects(X,Z) &
X: LargeTown, Y: Road, Z: Road

Q_3 = q(X) ← intersects(X,Y), adjacent_to(X,Z) &
X: LargeTown, Y: Road, Z: Water

Q_4 = q(X) ← intersects(X,Y) &
X: LargeTown, Y: Motorway

Q_5 = q(X) ← intersects(X,Y), intersects(X,Z) &
X: LargeTown, Y: Motorway, Z: MainTrunkRoad

Q_6 = q(X) ← intersects(X,Y), adjacent_to(X,Z) &
X: LargeTown, Y: Motorway, Z: Sea

Hierarchical Representations

- Research need:
  - The problem of generating irrelevant spatial patterns

    is_a(X,gasStation) → intersects(X,street)

    is amplified due to the hierarchies defined on spatial objects and spatial relations.

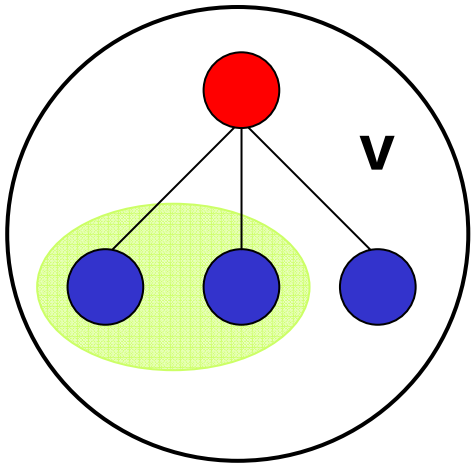
    is_a(X,fillingStation) → intersects(X,street)
    is_a(X,ShellStation) → intersects(X,street)
    is_a(X,fillingStation) → intersects(X,highway)
    is_a(X,fillingStation) → relateTo(X,highway)

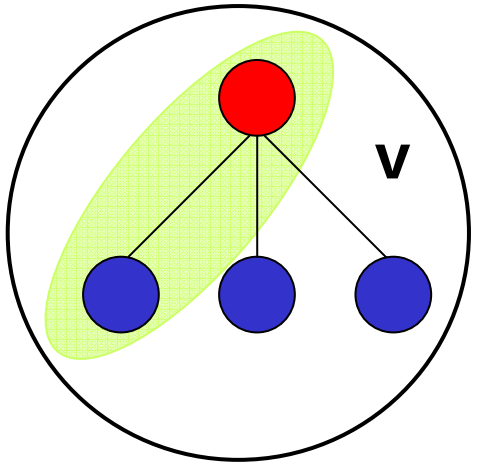
    …

  No obvious generalization of current techniques.

Dealing with Spatial Correlation

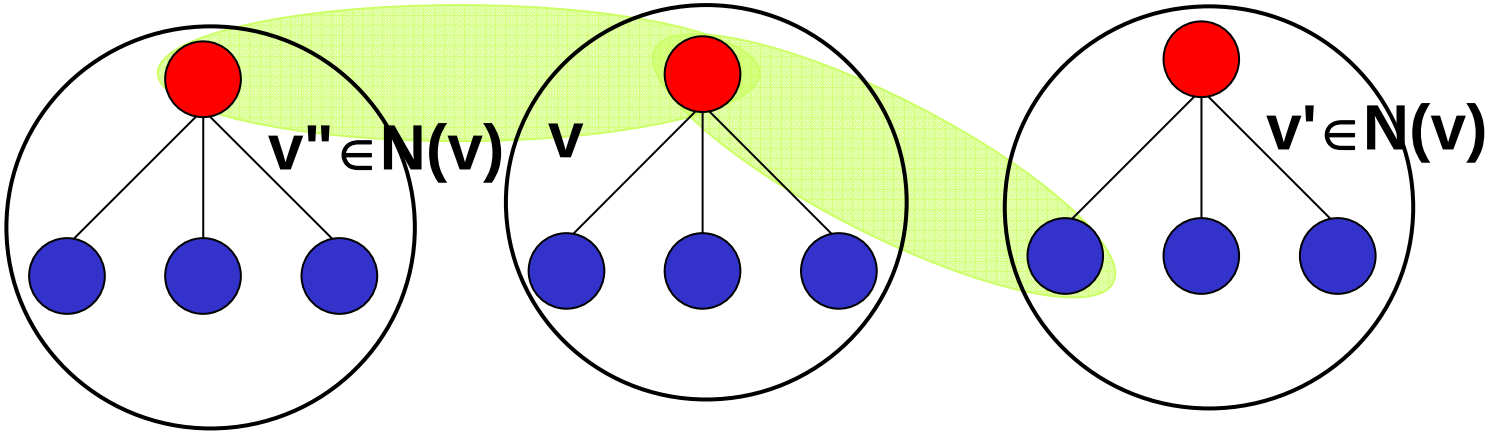
- Relational mining algorithms exploit two sources of correlation:
  - Local correlation, between attributes of each unit of analysis





Dealing with Spatial Correlation

- Relational mining algorithms exploit two sources of correlation:
  - Within-network correlation, between attributes of various units of analysis



Dealing with Spatial Correlation

- Some spatial data mining systems are based on general-purpose relational mining algorithms.
  - SubgroupMiner extends Wrobel's work on multirelational discovery of subgroups (PKDD'97)
  - MrsSMOTI extends Mr-SMOTI (ILP'03)

Dealing with Spatial Correlation

- Predictive modeling in spatial domains still challenges most relational mining algorithms when autocorrelation on the target attribute is captured.

$$y_i = f(x_i, x_{N(i)}, y_{N(i)})$$

Dependent variable in space i

Dependent variable in space N(i)

Both y_i and $y_{N(i)}$ have to be inferred collectively.

Dealing with Spatial Correlation

Collective inference outperforms independent classification when the autocorrelation between linked instances in the data graph is high.

D. Jensen, J., Neville, B., Gallagher, B. (2004): Why collective inference improves relational classification. KDD

Dealing with Spatial Correlation

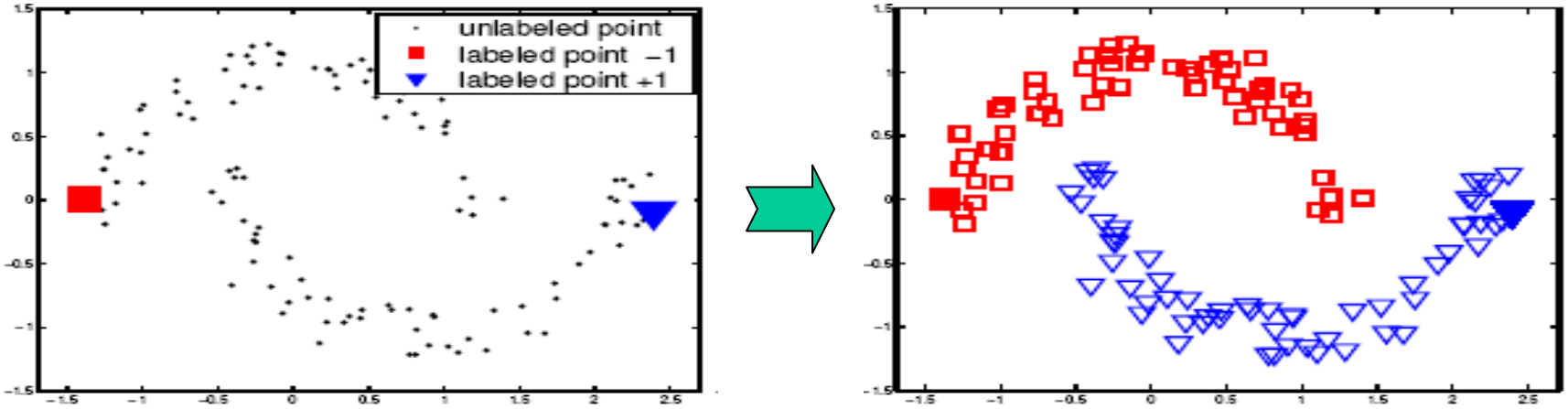
Autocorrelation is common in spatial domains.

Tobler's firt law of Geography

- Research need
  - Is collective inference useful in spatial predictive modeling tasks (classification, co-location, regression, …) ?

Dealing with Unlabelled Data

- Manual annotation on a map can be very demanding.
- Only few labels available
- Can we benefit of unlabelled examples?



Dealing with Unlabelled Data

- Two learning settings:
  - Transductive learning takes a "closed-world" assumption, i.e., the test data set is known in advance and the goal of learning is to optimize the generalization ability on this test data set, while the unlabeled examples are exactly the test examples.
  - Semi-supervised learning takes an "open-world" assumption, i.e., the test data set is not known and the unlabeled examples are not necessary test examples.

Dealing with Unlabelled Data

- Both are based on the semi-supervised smoothness assumption

  If two points x_1 and x_2 in a high-density region are close, then so should be the corresponding outputs y_1, y_2.
- The label function is smoother in high-density regions than in low-density regions.
- If two points are separated by a low-density region, then their outputs need not be close.

Dealing with Unlabelled Data

- If closeness of points correspond to some spatial distance measure, positive spatial autocorrelation entails the semisupervised smoothness assumption.

Research need:

Can a strong spatial autocorrelation counterbalance the lack of labelled data when learning in one of these settings?

Dealing with Unlabelled Data

Recent works on (relational) transductive learning in spatial domain have a positive reply …

M. Ceci, A., Appice, D., Malerba (2010): Transductive learning for spatial data classification. In: Koronacki, J., Ras, Z.W., Wierzchon, S.T., Kacprzyk, J. (eds.) Advances in Machine Learning I. Studies in Computational Intelligence, vol. 262, pp. 189–207. Springer

A. Appice, M., Ceci, D., Malerba (2010): Transductive learning for spatial regression with co-training. ACM SAC

Adding the Temporal Dimension

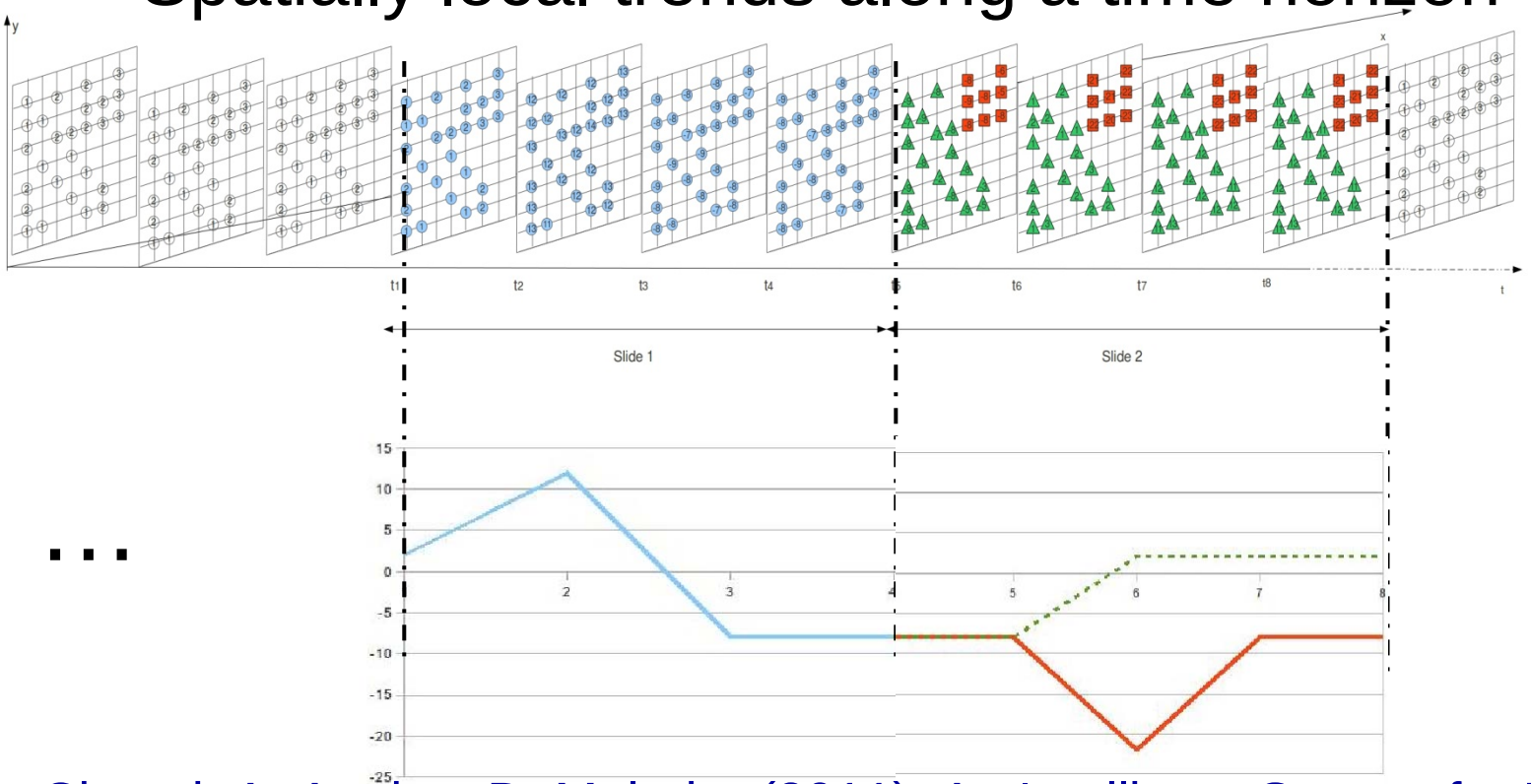
- Increasing trend: spatial data are augmented with a temporal dimension
- Changes may occur in time at the level
  - One or more attributes of a spatial objects

  Georeferenced sensors
  - The topology of the network of spatial objects

  Moving objects
- Time makes objects more complex (structured) and/or adds temporal relationships.

Adding the Temporal Dimension

Spatially local trends along a time horizon



A. Ciampi, A. Appice, D. Malerba (2011): An Intelligent System for Real Time Fault Detection in PV Plants. Proc. Sustainability in Energy and Buildings

Summary

- Mining spatial data presents several peculiarities
  - Spatial objects are heterogeneous and implicitly related by (infinitely) many spatial relationships;
  - Various forms of statistical dependence to consider
- Solutions offered in spatial statistics are limited
  - Double-entry table representation
  - The choice of neighborhood matrix is critical
  - Spatial dependencies handled in pre-processing

Summary

- The relational approach seems more appropriate
  - Several methods have already been proposed for different tasks
- But, there are still many challenges
  - Dynamic handling of spatial dependencies & scalability
  - Dealing with correlation and autocorrelation
  - Collective inference
  - Exploiting unlabelled data
  - Hierarchies of objects
  - Mining changes in space and time
  - …

Summary

- To develop effective solutions to spatial data analysis it is necessary to develop synergies between researchers working on different research areas:
  - Spatial statistics
  - Geocomputation & Geovisualization
  - Relational learning
  - Spatial Databases and GIS

- Motivation for optimism: real applications (e.g., sales prediction of individual shops, urban data analysis, location based services) demand for this collaboration.

Acknowledgements

Spatio-Temporal Data Mining @ Department of Computer Science. University of Bari “Aldo Moro”

- Annalisa Appice
- Michelangelo Ceci
- Anna Ciampi
- Antonietta Lanza
- Corrado Loglisci