

Analyzing Multi-level Spatial Association Rules Through a Graph-Based Visualization

Annalisa Appice and Paolo Buono

Dipartimento di Informatica,
Università degli Studi di Bari, Italy
{appice, buono}@di.uniba.it

Abstract. Association rules discovery is a fundamental task in spatial data mining where data are naturally described at multiple levels of granularity. ARES is a spatial data mining system that takes advantage from this taxonomic knowledge on spatial data to mine multi-level spatial association rules. A large amount of rules is typically discovered even from small set of spatial data. In this paper we present a graph-based visualization that supports data miners in the analysis of multi-level spatial association rules discovered by ARES and takes advantage from hierarchies describing the same spatial object at multiple levels of granularity. An application on real-world spatial data is reported. Results show that the use of the proposed visualization technique is beneficial.

1 Introduction

The rapidly expanding amount of spatial data gathered by collection tools, such as satellite systems or remote sensing systems has paved the way for advances in spatial databases. A spatial database contains (spatial) objects that are characterized by a geometrical representation (e.g. point, line, and region in a 2D context), a relative positioning with respect to some reference system as well as several non-spatial attributes. The widespread use of spatial databases in real-world applications, ranging from geo-marketing to environmental analysis or planning, is leading to an increasing interest in spatial data mining, i.e. extracting interesting and useful knowledge not explicitly stored in spatial databases.

Spatial association rules discovery is an important task of spatial data mining that aims at discovering interactions between reference objects (i.e. units of observation in the analysis) and one or more spatially referenced target-relevant objects or space dependent attributes, according to a particular spacing or set of arrangements. This task presents two main sources of complexity that is the implicit definition of spatial relations and the granularity of the spatial objects. The former is due to geometrical representation and relative positioning of spatial objects which implicitly define spatial relations of different nature, such as directional and topological. The second source of complexity refers to the possibility of describing the same spatial object at multiple levels of granularity. For instance, United Kingdom census data can be geo-referenced with respect to the hierarchy of areal objects ED \rightarrow Ward \rightarrow District \rightarrow County, based on the internal relationship between locations. This suggests that taxonomic knowledge

on task-relevant objects may be taken into account to obtain multi-level spatial association rules (descriptions at different granularity levels).

A full-fledged system that copes with both these issues is ARES (Association Rules Extractor from Spatial data) [1] that integrates SPADA (Spatial Pattern Discovery Algorithm) [6] to extract multi-level spatial association rules by exploiting an Inductive Logic Programming (ILP) approach to (multi-) relational data mining [5]. ARES assists data miners in extracting the units of analysis (i.e. reference objects and task-relevant objects) from a spatial database by means of a complex data transformation process that makes spatial relations explicit, and generates high-level logic descriptions of spatial data by specifying the background knowledge on the application domain (e.g. hierarchies on target-relevant spatial objects or knowledge domain) and defining some form of search bias to filter only association rules that fulfill user expectations.

Nevertheless, ARES may produce thousands of multi-level spatial association rules that discourage data miners to manually inspect them and pick those rules that represent true nuggets of knowledge at different granularity levels. A solution can be found in the emerging field of visual data mining that combines achievement of data mining with visual representation techniques leading to discovery tools that enable effective data (pattern) navigation and interpretation, preserve user control, and provide the possibility to discover anything interesting or unusual without the need to know in advance what kind of phenomena should be observed [4].

While a lot of research has been conducted on designing association rules exploratory visualization [3], no work, in our knowledge, properly deal with multi-level spatial association rules. At this aim, we propose to extend the graph-based visualization presented in [2] in order to visualize, navigate and interpret multi-level spatial association rules by exploiting both the knowledge embedded on hierarchies describing the same spatial object at multiple levels of granularity and the number of refinement steps performed to generate each rule.

The paper is organized as follows. The problem of mining multi-level spatial association rules with ARES is discussed in Section 2, while the graph-based approach to visualize multi-level spatial association rules is presented in Section 3. An application to mine North West England 1998 census data is then discussed in Section 4. The goal is to investigate the mortality rate according to both socio-economic deprivation factors represented in census data and geographical factors represented in topographic maps. At this aim, multi-level spatial association rules discovery is combined with a graph-based visualization to pick those rules which may provide a guidance to recognize and balance the multiple factors affecting the mortality risk. Finally, conclusions are drawn.

2 Multi-level Spatial Association Rules Mined with ARES

The problem of mining multi-level spatial association rules can be formally defined as follows: *Given* a spatial database (SDB), a set S of reference objects,

some sets R_k , $1 \leq k \leq m$, of task-relevant objects, a background knowledge BK including some spatial hierarchies H_k on objects in R_k , M granularity levels in the descriptions (1 is the highest while M is the lowest), a set of granularity assignments ψ_k which associate each object in H_k with a granularity level, a couple of thresholds $\text{minsup}[l]$ and $\text{minconf}[l]$ for each granularity level, a language bias LB that constrains the search space; *Find* strong multi-level spatial association rules, that is, association rules involving spatial objects at different granularity levels.

The reference objects are the main subject of the description, namely units of observation, while the task-relevant objects are spatial objects that are relevant for the task in hand and are spatially related to the former. Both the set of target object S and the sets of target relevant objects R_k typically correspond with layers of the spatial database, while hierarchies H_k define *is-a* (i.e., taxonomical) relations of spatial objects in the same layer (e.g. regional road *is-a* road, main trunk road *is-a* road, road *is-a* transport net). Objects of each hierarchy are mapped to one or more of the M user-defined description granularity levels in order to deal uniformly with several hierarchies at once. Both frequency of patterns and strength of rules depend on the granularity level l at which patterns/rules describe data. Therefore, a pattern P ($s\%$) at level l is frequent if $s \geq \text{minsup}[l]$ and all ancestors of P with respect to H_k are frequent at their corresponding levels. The support s estimates the probability $p(P)$. An association rule $A \rightarrow C$ ($s\%$, $c\%$) at level l is strong if the pattern $A \cup C$ ($s\%$) is frequent and $c \geq \text{minconf}[l]$, where the confidence c , estimates the probability $p(C|A)$ and A (C) represents the antecedent (consequent) of the rule.

Since a spatial association rules is an association rule whose corresponding pattern is spatial (i.e. it captures a spatial relationship among a spatial reference object and one or more target-relevant spatial object or space dependent attributes), it can be expressed by means of predicate calculus. An example of spatial association rule is "*is_a*(X , *town*), *intersects*(X , Y), *is_a*(Y , *road*) \rightarrow *intersects*(X , Z), *is_a*(Z , *road*), $Z \neq Y$ (91%, 100%)" to be read as "if a town X intersects a road Y then X intersects a road Z distinct from Y with 91% support and 100% confidence", where X denotes a target object in town layer, while Y and Z some target-relevant object in road layer. By taking into account taxonomic knowledge on task-relevant objects in the road layer, it is possible to obtain descriptions at different granularity levels (multiple-level spatial association rules). For instance, a finer-grained association rules can be "*is_a*(X , *town*), *intersects*(X , Y), *is_a*(Y , *regional_road*) \rightarrow *intersects*(X , Z), *is_a*(Z , *main_trunk_road*), $Z \neq Y$ (65%,71%)", which states that "if a town X intersects a regional road Y then X intersects a main trunk road Z distinct from Y with 65% support and 71% confidence."

The problem above is solved by the algorithm SPADA that operates in three steps for each granularity level: i) pattern generation; ii) pattern evaluation; iii) rule generation and evaluation. SPADA takes advantage of statistics computed at granularity level l when computing the supports of patterns at granularity level $l + 1$.

In the system ARES¹, SPADA has been loosely coupled with a spatial database, since data stored in the SDB Oracle Spatial are pre-processed and then represented in a deductive database (DDB). Therefore, a middle layer is required to make possible a loose coupling between SPADA and the SDB by generating features of spatial objects. This middle layer includes both the module RUDE (Relative Unsupervised Discretization) to discretize a numerical attribute of a relational database in the context defined by other attributes [7] and the module FEATEX (Feature Extractor) that is implemented as an Oracle package of procedures and functions, each of which computes a different feature. According to their nature, features extracted by FEATEX can be distinguished as geometrical (e.g. area and length), directional (e.g. direction) and topological features (e.g. crosses) [1]. Extracted features are then represented by extensional predicates. For instance, spatial intersection between two objects X and Y is expressed with $crosses(X, Y)$. In this way, the expressive power of first-order logic in databases is exploited to specify both the background knowledge BK , such as spatial hierarchies and domain specific knowledge, and the language bias LB . Spatial hierarchies allow to face with one of the main issues of spatial data mining, that is, the representation and management of spatial objects at different levels of granularity, while the domain specific knowledge stored as a set of rules in the intensional part of the DDB supports qualitative spatial reasoning. On the other hand, the LB is relevant to allow data miners to specify his/her bias for interesting solutions, and then to exploit this bias to improve both the efficiency of the mining process and the quality of the discovered rules. In SPADA, the language bias is expressed as a set of constraint specifications for either patterns or association rules. Pattern constraints allow to specify a literal or a set of literals that should occur one or more times in discovered patterns. During the rule generation phase, patterns that do not satisfy a pattern constraint are filtered out. Similarly, rule constraints are used to specify literals that should occur in the head or body of discovered rules.

In a more recent release of SPADA (3.1) new pattern (rule) constraints have been introduced in order to specify exactly both the minimum and maximum number of occurrences for a literal in a pattern (head or body of a rule). Moreover, an additional rule constraint has been introduced to eventually specify the maximum number of literals to be included in the head of a rule. In this way users may define the head structure of a rule requiring the presence of exactly a specific literal and nothing more. In this case, the multi-level spatial association rules discovered by ARES may be used for sub-group discovery tasks.

3 Multi-level Spatial Association Rules Graph-Based Visualization

A set R of multi-level spatial association rules can be naturally partitioned into $M \times N$ groups denoted by R_{ij} , where i ($1 \leq i \leq M$) denotes the level of

¹ <http://www.di.uniba.it/malerba/software/ARES/index.htm>

granularity in the spatial hierarchies H_k , while j ($2 \leq j \leq N$) the number of refinement steps performed to obtain the pattern (i.e. number of atoms in the pattern). Each set R_{ij} can be visualized in form of a graph by representing antecedent and consequent of rules as nodes and relationships among them as edges.

This graph-based visualization can be formally defined as follows: Given an association rules set R , a directed (not completely connected) graph $G = (N, E)$ can be built from R , such that:

- N is a set of couples (l, t) , named *nodes*, where l denotes the conjunction of atoms representing the antecedent (A) or consequent (C) of a rule $A \rightarrow C \in R$, while t is a flag denoting the node role (i.e. antecedent, consequent or both of them).
- E is a set of 4-tuples (n_A, n_C, s, c) , named *edges*, where n_A is a node with the role of antecedent; n_C is a node with the role of consequent, while s and c are the support and confidence of the rule $n_A.l \rightarrow n_C.l \in R$ respectively.

Each node of G can be visualized as a colored circle: a red circle represents a node n with the role of antecedent ($n.t = antecedent$) while a green circle represents a node n with the role of consequent ($n.t = consequent$). If the node has the role of antecedent for a rule and consequent for a different rule, it appears half red and half green. The label $n.l$ can be visualized in a rectangular frame close to the circle representing n . Conversely, each edge in G can be visualized by a straight segment connecting the node n_A with the node n_C . It corresponds with the rule $n_A.l \rightarrow n_C.l$ that exists in R . The confidence of this rule is coded by the length of the edge, the greater is the confidence, the longer is the edge. Conversely, the support is coded by color saturation of the edge: from light blue (low support) to black (high support). Support and/or confidence can be also visualized in a text label close to the edge (see Figure 1).

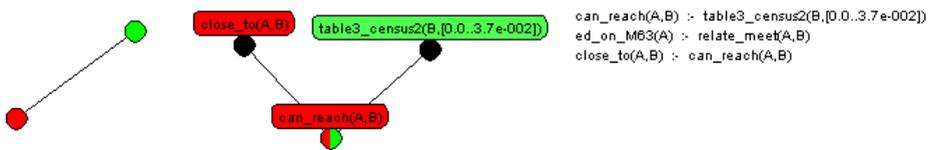


Fig. 1. Visualizing the graph of spatial association rules

As suggested by [2], this graph representation appears beneficial in exploring huge amount of association rules in order to pick interesting and useful patterns, since it takes advantages from human perceptual and cognitive capabilities to immediately highlight which association rules share the same antecedent or consequent with respect to the overall distribution of rules. Filtering mechanisms which permit to hide a sub-graph of G (i.e. subset of rules in R) according to either minimal values of support and confidence or the absence of one or more predicates in the rule provide a better interaction.

To explore multi-level spatial association rules discovered by ARES, this graph-based visualization should be further extended in order to enable data miners to navigate among several graphs G_{ij} according to either the levels of granularity i or the number of refinement steps j . In the former case, for each pair of granularity levels (i, h) with $1 \leq i < h \leq M$ ($1 \leq h < i \leq M$) and number of refinement steps j ($2 \leq j \leq N$), a specialization (generalization) operator $\rho_{i \downarrow h, j}$ ($\delta_{i \uparrow h, j}$) can be defined as follows:

$$\rho_{i \downarrow h, j} : R_{ij} \rightarrow \wp(R_{hj}) \quad (\delta_{i \uparrow h, j} : R_{ij} \rightarrow \wp(R_{hj})),$$

where $\wp(R_{hj})$ denotes the power set of R_{hj} . For each spatial association rule $A \rightarrow C \in R_{ij}$, $\rho_{i \downarrow h, j}(A \rightarrow C) = \{A_1 \rightarrow C_1, \dots, A_w \rightarrow C_w\}$, such that each $A_k \rightarrow C_k \in R_{hj}$ ($k = 1, \dots, w$) and $A_k \rightarrow C_k$ is a down-specialization (up-generalization) of $A \rightarrow C$.

To formally define the relation of down-specialization (up-generalization) between two spatial association rules, we represent each spatial rule $A \rightarrow C$ as $A_S, A_I \rightarrow C_S, C_I$, where A_S (C_S) includes all atoms in A (C) describing either a property (e.g. *area*(X , [10..15]) or *cars*(X , [150..1000])), a relationship (e.g. *intersect*(X , Y)) or an inequality (e.g. $X/ = Y$). Conversely, A_I (C_I) includes all *is_a* atoms (e.g. *is_a*(X , *road*)). Therefore, $A' \rightarrow C' \in R_{hj}$ is a down-specialization of $A \rightarrow C \in R_{ij}$ iff there exists a substitution θ (i.e. a function that associates a variable with a term) that renames variables in $A' \rightarrow C'$ such that $A_S = A'_S\theta$, $C_S = C'_S\theta$, and for each *is_a* atom of A_I (C_I) in the form *is_a*(X , v_i), where X denotes a target relevant object in R_k and v_i is a node at level i of the spatial hierarchy H_k , there exists an atom *is_a*(X , v_h) in $A'_I\theta$ ($C'_I\theta$) with v_h a node in the sub-hierarchy of H_k that is rooted in v_i . The up-generalization differs from down-specialization only in requiring that v_i is a node in the sub-hierarchy of H_k that is rooted in v_h and not vice-versa.

Example: Let us consider the spatial association rules:

R1: *intersects*($X1$, $Y1$), *cars*($X1$, [25, 120]), *is_a*($X1$, *town*), *is_a*($Y1$, *road*)
 \rightarrow *mortality*($X1$, *high*).

R2: *intersects*($X2$, $Y2$), *cars*($X2$, [25, 120]), *is_a*($X2$, *town*),
is_a($Y2$, *main_trunk_road*) \rightarrow *mortality*($X2$, *high*).

where $R1.A_S$ is "*intersects*($X1$, $Y1$), *cars*($X1$, [25, 120])" and $R1.C_S$ is "*mortality*($X1$, *high*)", while $R1.A_I$ is "*is_a*($X1$, *town*), *is_a*($Y1$, *road*)" and $R1.C_I$ is empty. Similarly $R2.A_S$ is "*intersects*($X2$, $Y2$), *cars*($X2$, [25, 120])" and $R2.C_S$ is "*mortality*($X2$, *high*)", while $R2.A_I$ is "*is_a*($X2$, *town*), *is_a*($Y2$, *main_trunk_road*)" and $R2.C_I$ is empty. R2 is a *down-specialization* of R1 since there exists the substitution $\theta = \{X2/X1, Y2/Y1\}$ such that $R1.A_S = R2.A_S\theta$, $R1.C_S = R2.C_S\theta$, and *main_trunk_road* is a specialization of *road* in the corresponding hierarchy. Conversely, R1 is an *up-generalization* of R2.

A different specialization (generalization) operator $\rho_{i, j \rightarrow h}$ ($\delta_{i, j \leftarrow h}$) can be further defined, for each granularity level i and pair of refinement step numbers (j, h) with $2 \leq j < h \leq N$ ($2 \leq h < j \leq N$), such that:

$$\rho_{i, j \rightarrow h} : R_{ij} \rightarrow \wp(R_{ih}) \quad (\delta_{i, j \leftarrow h} : R_{ij} \rightarrow \wp(R_{ih})),$$

In this case, for each spatial association rule $A \rightarrow C \in R_{ij}$, $\rho_{i,j \rightarrow h}(A \rightarrow C) = \{A_1 \rightarrow C_1, \dots, A_w \rightarrow C_w\}$, where $A_k \rightarrow C_k \in R_{ih}$ ($k = 1, \dots, w$) and $A_k \rightarrow C_k$ is a right-specialization (left-generalization) of $A \rightarrow C$. More formally, a spatial association rule $A' \rightarrow C' \in R_{ih}$ is a right-specialization (left-generalization) of $A \rightarrow C \in R_{ij}$ iff there exists a substitution θ such that $A\theta \subset A'$ and $C\theta \subset C'$ ($A'\theta \subset A$ and $C'\theta \subset C$).

Example: Let us consider the spatial association rules:

R1: `is_a(X1, town), intersects(X1, Y1), is_a(Y1, road) → mortality(X1, high)`
 R2: `is_a(X1, town), intersects(X1, Y1), is_a(Y1, road), extension(Y1, [12..25])`
`→ mortality(X1, high)`.

R2 is a *right-specialization* of R1, since there exists the substitution $\theta = \{X1/X2, Y1/Y2\}$ such that $R1.A\theta \subset R2.A$ and $R1.C\theta \subset R2.C$. Conversely, R1 is a *left-generalization* of R2.

Consequently, by combining a multiple graph visualization with operators of both specialization and generalization defined above, data miners are able to navigate among the graphs G_{ij} . This means that it is possible to down(right)-specialize or up(left)-generalize the portion of the graph G_{ij} representing a specific rule $R \in R_{ij}$ and visualize the corresponding sub-graph of spatial association rules extracted at a different level of granularity or number of refinement steps.

This graph-based visualization has been implemented into a visualization tool, named ARVis (multi-level Association Rules Visualizer), which actively supports data miners in exploring and navigating among several graphs of multi-level association rules G_{ij} by highlighting the portion of graph that represents the down (right)-specialization or up(left)-generalization of a rule, zooming rules, dynamically filtering rules according to minimal values of support and/or confidence as well as presence or absence of some relevant predicate and visualizing details about a rule (e.g. support, confidence, patterns, rules).

4 An Application: Mining Geo-Referenced Data

In this section we present a real-world application concerning with both mining and exploring multi-level spatial association rules for geo-referenced census data interpretation. We consider census and digital map data stored into an Oracle Spatial 9i database provided in the context of the European project SPIN! (Spatial Mining for Data of Public Interest) [8]. This data concerns Greater Manchester, one of the five counties of North West England, which is divided into censual sections or wards, for a total of 214 wards. Spatial analysis is enabled by the availability of vectorized boundaries of the 1998 greater Manchester census wards as well as Ordnance Survey digital maps where several interesting layers are found (e.g. urban area or road net). Census data, geo-referenced at ward level, provide socio-economic statistics (e.g. mortality rate that is the percentage of deaths with respect to the number of inhabitants) as well as some measures describing the deprivation level (e.g. Townsend index, Carstairs index, Jarman index and DoE index). Both mortality rate and deprivation indices are

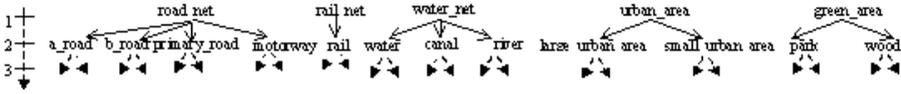


Fig. 2. Spatial hierarchies defined for five Greater Manchester layers: road net, rail net, water net, urban area and green area

all numeric. They can be automatically discretized with ARES. More precisely, Jarman index, Townsend index, DoE index and Mortality rate are automatically discretized in (low, high), while Carstairs index is discretized in (low, medium, high).

For this application, we decide to employ ARES in mining multi-level spatial association rules relating Greater Manchester wards, which play the role of reference object, with topological related roads, rails, waters, green areas and urban areas as task relevant objects. We extract 784,107 facts concerning topological relationships between each relevant object and task relevant object stored in the spatial database for Greater Manchester area. An example of fact extracted is *crosses(ward_135, urbareaL_151)*. However, to support a spatial qualitative reasoning, we also express a domain specific knowledge (*BK*) in form of a set of rules. Some of these rules are:

$$\begin{aligned} \text{crossed_by_urbanarea}(X, Y) &: - \text{crosses}(X, Y), \text{is_a}(Y, \text{urban_area}). \\ \text{crossed_by_urbanarea}(X, Y) &:- \text{inside}(X, Y), \text{is_a}(Y, \text{urban_area}). \end{aligned}$$

Here the use of the predicate *is_a* hides the fact that a hierarchy has been defined for spatial objects which belong to the urban area layer. In detail, five different hierarchies are defined to describe the following layers: road net, rail net, water net, urban area and green area (see Figure 2). The hierarchies have depth three and are straightforwardly mapped into three granularity levels. They are also part of the *BK*. To complete the problem statement, we specify a language bias (*LB*) both to constrain the search space and to filter out uninteresting spatial association rules. We rule out all spatial relations (e.g. *crosses*, *inside*, and so on) directly extracted from spatial database and ask for rules containing topological predicates defined by means of *BK*. Moreover, by combining the rule filters *head_constraint([mortality_rate(-)], 1, 1)* and *rule_head_length(1, 1)* we ask for rules containing only mortality rate in the head. In addition, we specify the maximum number of refinement steps as $J = 8$ and the minimal values of support and confidence for each granularity level as: $\text{minsup}[1]=0.1$, $\text{minsup}[2]=0.1$, $\text{minsup}[3]=0.05$, $\text{minconf}[1]=0.3$, $\text{minconf}[2]=0.2$ and $\text{minconf}[3]=0.1$.

ARES generates 239 strong rules at first granularity level, 1140 at second granularity level and 15 at third granularity level. These rules are extracted from a set of 28496 frequent patterns describing the geographically distributed phenomenon of mortality in Greater Manchester at different granularity levels with respect to the spatial hierarchies we have defined on road, rail, water, urban area and green area layers. To explore this huge amount of multi-level spatial association rules and find which rules can be a valuable support to good public policy, we exploit the multiple graph-based visualization implemented in ARVis.

In this way, we are able to navigate among different graphs G_{ij} ($i = 1, \dots, 3$ and $j = 2, \dots, 8$) representing the group of rules R_{ij} discovered by ARES at i granularity level after j refinement steps. For instance, Figure 3 shows the graph of spatial association rules G_{15} . By graphically filtering rules in G_{15} according to confidence value, we identify the most confident rule $R1$ that is: $is_a(A, ward), crossed_by_urbanarea(A, B), is_a(B, urban_area), townsendidx_rate(A, high) \rightarrow mortality_rate(A, high)$ ($c=39.71\%, s=70.24\%$). This rule states that a high mortality rate is observed in a ward A that includes an urban area B and has a high value of Townsend index. The support (39.71%) and the high confidence (70.24%) confirm a meaningful association between a geographical factor such as living in deprived urban areas and a social factor such as the mortality rate. The same rule is highlighted in the graph G_{15} by filtering with respect

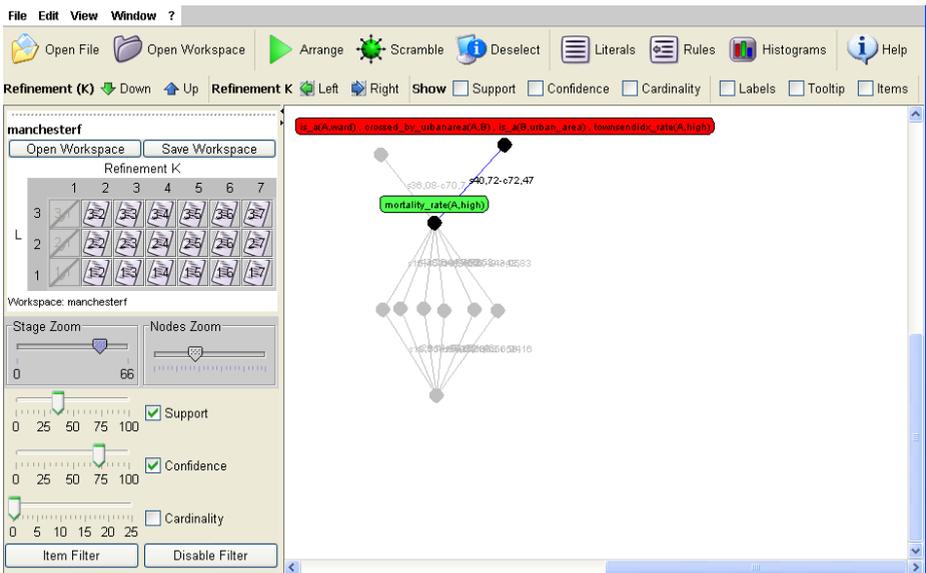


Fig. 3. Visualizing the graph of spatial association rules using ARVis

to increasing value of support. Moreover, by left-generalizing $R1$, we navigate from the graph G_{15} to a portion of the graph G_{14} and identify the rule $R2$ that is $is_a(A, ward), crossed_by_urbanarea(A, B), is_a(B, urban_area) \rightarrow mortality_rate(A, high)$ (54.67%, 60.3%). This rule has a greater support and a lower confidence. The same rule is highlighted in the entire graph G_{14} by graphically filtering with respect to increasing values of support and confidence. These two association rules show together an unexpected association between Townsend index and urban areas. Apparently, this means that this deprivation index is unsuitable for rural areas.

Conversely, we may decide to up-generalize $R1$ and move from the graph G_{15} to the portion of the graph G_{25} representing association rules which are up-generalization of $R2$ mined by ARES at second granularity level after four refinement steps. In this way, we discover that, at second granularity level, SPADA specializes the task relevant object B by generating the following rule which preserve both support and confidence: $R3: is_a(A, ward), crossed_by_urbanarea(A, B), is_a(B, urban_areaL), townsendidx_rate(A, high) \rightarrow mortality_rate(A, high)$ (39.71%, 70.24%). This rule clarifies that the urban area B is large. Similar considerations are suggested when we explore graphs of multi-level spatial association rules generated after more refinement steps.

We may explore spatial association rules characterizing low mortality wards. By visualizing G_{15} and moving the confidence filter slider, we discover that the highest confident rule with low mortality in the consequent is: $is_a(A, ward), crossed_by_urbanarea(A, B), is_a(B, urban_area), townsendidx_rate(A, low) \rightarrow mortality_rate(A, low)$ (19.15%, 56.16%), stating that a low valued Townsend index ward A that (partly) includes an urban area B presents a low mortality.

5 Conclusions

In this paper we have presented a graph-based visualization specially designed to support data miners in exploring multi-level spatial association rules and finding true nuggets of knowledge. This new visualization extend traditional graph-based technique with operators of both generalization and specialization that allow data miners to navigate among different graphs of spatial association rules partitioned according with both the granularity level in spatial hierarchies and the number of refinement steps in generating the corresponding pattern. A real-world application shows that this visualization is beneficial for exploring multi-level spatial association rules discovered by ARES. Currently, usability testing are going on, and results will be provided in a future work.

References

1. A. Appice, M. Ceci, A. Lanza, F. A. Lisi, and D. Malerba. Discovery of spatial association rules in georeferenced census data: A relational mining approach, intelligent data analysis. *Intelligent Data Analysis*, 7(6):541–566, 2003.
2. D. Bruzzese and P. Buono. Combining visual techniques for association rules exploration. In M. F. Costabile, editor, *Proceedings of the Working Conference on Advanced Visual Interfaces AVI 2004*, pages 381–384. ACM Press, 2004.
3. D. Bruzzese and C. Davino. Visual post analysis of association rules. *Journal of Visual Languages and Computing*, 14(6):621–635, 2003.
4. M. F. Costabile and D. Malerba. Special issue on visual data mining, editor’s foreword. *Journal of Visual Languages & Computing*, 14(6):499–501, 2003.
5. S. Džeroski and N. Lavrač, editors. *Relational Data Mining*. Springer, 2001.
6. F. Lisi and D. Malerba. Inducing multi-level association rules from multiple relations. *Machine Learning*, 55:175–210, 2004.

7. M. Ludl and G. Widmer. Relative unsupervised discretization for association rule mining. In D. Zighed, H. Komorowski, and J. Zytkow, editors, *Principles of Data Mining and Knowledge Discovery*, volume 1910 of *LNAI*, pages 148–158. Springer-Verlag, 2000.
8. M. May. Spatial knowledge discovery: The SPIN! system. In K. Fullerton, editor, *Proceedings of the EC-GIS Workshop*, 2000.