A general framework for trajectory data warehousing and visual OLAP

Luca Leonardi, Salvatore Orlando, Alessandra Raffaetà, Alessandro Roncato, Claudio Silvestri, Gennady Andrienko & Natalia Andrienko

GeoInformatica
An International Journal on Advances of Computer Science for Geographic Information Systems

ISSN 1384-6175

Geoinformatica
DOI 10.1007/s10707-013-0181-3

Springer

Geoinformatica
DOI 10.1007/s10707-013-0181-3

A general framework for trajectory data warehousing and visual OLAP

Luca Leonardi · Salvatore Orlando · Alessandra Raffaetà · Alessandro Roncato · Claudio Silvestri · Gennady Andrienko · Natalia Andrienko

Received: 17 October 2012 / Revised: 4 March 2013 / Accepted: 21 April 2013

Abstract In this paper we present a formal framework for modelling a trajectory data warehouse (TDW), namely a data warehouse aimed at storing aggregate information on trajectories of moving objects, which also offers visual OLAP operations for data analysis. The data warehouse model includes both temporal and spatial dimensions, and it is flexible and general enough to deal with objects that are either completely free or constrained in their movements (e.g., they move along a road network). In particular, the spatial dimension and the associated concept hierarchy reflect the structure of the environment in which the objects travel. Moreover, we cope with some issues related to the efficient computation of aggregate measures, as needed for implementing roll-up operations. The TDW and its visual interface allow one to investigate the behaviour of objects inside a given area as well as the movements of objects between areas in the same neighbourhood. A user can

L. Leonardi · S. Orlando · A. Raffaetà · A. Roncato · C. Silvestri (✉)
DAIS, Università Ca' Foscari Venezia, Venezia, Italy
e-mail: silvestri@unive.it

L. Leonardi
e-mail: leonardi@dsi.unive.it

S. Orlando
e-mail: orlando@dsi.unive.it

A. Raffaetà
e-mail: raffaeta@dsi.unive.it

A. Roncato
e-mail: roncato@dsi.unive.it

G. Andrienko · N. Andrienko
IAIS, Fraunhofer Institute, Sankt Augustin, Germany

G. Andrienko
e-mail: gennady.andrienko@iais.fraunhofer.de

N. Andrienko
e-mail: natalia.andrienko@iais.fraunhofer.de

Published online: 11 May 2013

easily navigate the aggregate measures obtained from OLAP queries at different granularities, and get overall views in time and in space of the measures, as well as a focused view on specific measures, spatial areas, or temporal intervals. We discuss two application scenarios of our TDW, namely road traffic and vessel movement analysis, for which we built prototype systems. They mainly differ in the kind of information available for the moving objects under observation and their movement constraints.

Keywords Spatio-temporal data warehouses · Visual analytics · Distinct count problem

1 Introduction

Nowadays the advances in mobile networks and devices, sensors, and space positioning technologies allow moving objects to be continuously tracked, thus producing huge streams of observations which can be used to reconstruct object trajectories. The analysis of trajectory data imposes new challenges for their efficient management, but also raises opportunities for discovering behavioral mobility patterns that can be exploited in many innovative applications, such as traffic management, service accessibility and monitoring, and control surveillance. In this context we argue that *data warehouse* technologies, combined with *geographic visual analytics* tools, can play an important role in granting very fast, accurate and understandable analysis of mobility data.

This paper proposes a general framework for modelling and implementing a *Trajectory Data Warehouse* (TDW), which offers visual OLAP operations for the analysis of aggregate data. The framework relies on a flexible *conceptual model* with associated *spatio-temporal dimensions* and *hierarchies*. More specifically, the spatial domain can be structured according to the application requirements, by exploiting hierarchies of regular grids (like in [36, 37]) or of regions with ad-hoc shapes. While a hierarchy of regular grids can be used to analyse objects that can move freely in the space, hierarchies with ad-hoc shapes are useful for objects whose movements are constrained, such as objects that can only move along a road network (e.g., cars). Moreover, Voronoi tessellation can be employed in order to build hierarchies of regions based on the actual distribution of the points forming the trajectories. This form of partitioning turns out to be particularly suited for highlighting the directions of the trajectory movement.

The distinctive features of our proposal are a general *formalisation* of relevant concepts, measures, and aggregation functions, and a *map-based Visual OLAP* interface offered by the system.

Concerning the formal setting, starting from a rigorous definition of granularity (i.e., a spatio-temporal partition of the spatial and temporal domains of interest), we formally introduce the concept of *trajectory decomposition* according to a given granularity. This is fundamental to provide formal definitions of the measures we want to store in our TDW. The modelled measures represent *aggregate information*, e.g., the number of trajectories visiting/starting/ending in a granule or crossing a granule border, or the cumulative travelled distance and the average speed of the trajectories traversing a granule. They report some significant features of the sets of trajectories crossing the granules, whereas single trajectory details, like object

identifiers, are not kept in the TDW. There are, in general, good reasons for this design choice. Often individual data can be highly volatile and require huge memory space. More importantly, in some cases they cannot be stored due to legal or privacy issues, and anonymization might not suffice to guarantee privacy of tracked objects/people [49]. In addition, for common spatio-temporal applications, aggregate measures are typically much more relevant than information about individual moving objects [51].

Relying on the formally defined measures we introduce corresponding aggregate functions, which are proved to be algebraic or distributive [23]. Such functions are used to compute roll-up queries, by linearly combining sub-aggregates stored in the lower levels of the hierarchy. Among the others, we introduce a novel measure *visits*, denoted by $\mathscr{V}$, which counts the number of times the various trajectories visits a given granule. The aggregation function for measure $\mathscr{V}$ is proved to be algebraic. An element of interest for $\mathscr{V}$ is the fact that it provides an approximation of measure *presence*, denoted by $\mathscr{P}$, which counts the number of *distinct* trajectories occurring in a spatio-temporal granule. The aggregate function for $\mathscr{P}$ is *holistic* [23]: this kind of functions represents a big issue for data warehouse technology and various solutions have been proposed in the literature. We will show that our technique is competitive with respect to some common approaches facing the same problem.

The *map-based Visual OLAP* interface allows for multidimensional and interactive analysis, it integrates OLAP tools with visual analytics [7], and it permits to overcome the limits of the usual OLAP user interfaces. In fact, the table based representation commonly adopted by OLAP tools makes it very difficult for the user to grasp the relationships between areas in the same neighbourhood, the evolution of spatial measures in time, or the correlations of different measures. We believe that visualisation is crucial: it can be seen simultaneously as the output and end-product of a knowledge discovery cycle and the starting point for further, interactive and visual, analysis.

To demonstrate the flexibility, and potentialities of the proposed TDW model, we instantiate it to create two real TDW prototype systems, one concerning vessels sailing in the sea and the other dealing with cars moving along a road network. The main technologies exploited are Oracle™ 11 DBMS (with spatial extension), and V-Analytics [4], which is a Java-based visual analytics system addressing the specifics of spatial and temporal data.

It is worth observing that our TDW approach has some advantages with respect to an alternative, more classical approach based on a Moving Object Database (MOD) [27] coupled with a map-based visual interface. If the MOD stores all the historical raw trajectory data, we could guarantee, in principle, to answer similar queries as the one allowed by our TDW. However, this requires to exploit the very expensive MOD spatio-temporal join operators in order to determine which are the moving objects that satisfy each query, i.e., the object trajectories that cross all the involved spatio-temporal granules. Instead, the TDW makes the resolution of such queries simpler, faster, and definitively more *scalable*, since it has only to recombine the already pre-computed sub-aggregate measures, associated with the various spatio-temporal granules affected by each query. For example, this allows an analyst to quickly prepare an *animated map* that shows the temporal evolution of the aggregates regarding a certain area for a given time interval. It is worth noting that the OLAP query resolution phase becomes simpler and faster using our TDW because we have moved most of the burden to the pre-processing ETL

(extract, transform, load) phase of the TDW. The ETL phase is not the focus of this paper and we will give only some hints about the reconstruction of the trajectories from raw data. It should be clear that during ETL, we still deal with all the typical spatio-temporal expensive spatial join operations to determine which trajectories affect/cross our base spatio-temporal granules. The advantage is that ETL is carried out only once. Obviously there is a price to pay for the gain in efficiency. The main drawback of our approach, which is common to DW data modeling, is that it may sacrifice some flexibility, since only queries exactly matching the spatio-temporal granularities allowed by the TDW hierarchies can be answered in a precise way.

In summary, the main contributions of this paper are:

(i) a theoretical framework for designing a TDW, with specific measures and suitable functions to aggregate these measures along the TDW spatio-temporal dimensions and hierarchies;
(ii) the possibility of structuring the spatial domain according to the specifics of trajectory data (the spatial hierarchy is no longer restricted to consist of simple regular grids as in our previous work [36, 37]).
(iii) a set of spatial and temporal visualisation techniques, supporting OLAP analysis of movement data;
(iv) the development of two case studies having different features: cars moving along a road network and vessels sailing on the sea. The resulting TDW prototypes are an effective demonstration of the flexibility and of the potentialities of our approach.

The rest of the paper is organised as follows. Section 2 discusses the related work. In Section 3 we present some scenarios where our TDW can be profitably applied. Section 4 introduces the TDW conceptual model and gives examples of its instantiation for the application scenarios, by specifying hierarchies and measures. Section 5 formally defines the spatio-temporal hierarchies, some stored measures and the associated aggregation functions used in the illustrated scenarios. Then, in Section 6, we cope with the issues related to approximating the measure *Presence*. In Section 7 we describe the visual OLAP interface of our system by real use cases whereas in Section 8 we give some hints on the implementation of the presented prototypes. Section 9 draws conclusions. Finally, Appendix A contains the proofs of the propositions of Section 5.

2 Related work

Trajectory data warehouses are a relatively new research topic, with significant intersection with at least two extensively studied fields: moving object databases (MOD) and spatial data warehouses (SDW).

Research on MOD started more than a decade ago, building on previous research on spatial and temporal databases, and is still very active. There is a large amount of literature on this topic, starting with the works of Güting et al. [16] and Wolfson et al. [55]. The reader may refer to [27] for an in-depth discussion of MOD, and to [41] and [26] for examples of MOD implementations based, respectively, on a commercial DBMS and on a platform for database prototyping, both referring to the case of unconstrained movement. With a few years delay with respect to the corresponding results for unconstrained movement, several papers have defined

data models to represent movement in constrained spaces, and in particular in networks [25, 39]. Further, modern MOD support sophisticated spatio-temporal queries incorporating time-dependent and sequential predicates [46]. Integration of MOD in a visual analytics environment enables complex analysis scenarios [45]. Despite the efficiency that a MOD can achieve by using indexes, both with [30, 42, 43] or without [12] movement constraints, and techniques to reduce the size of data related to a trajectory while preserving error bounds [11], storing the whole moving object trajectories has severe consequences on response time and storage space requirements in case the amount of data to manage is unbounded.

In non-spatial contexts, it is common for analytical applications to make use of dedicated collections of subject-oriented, integrated, non-volatile, and time-variant data, referred to as data warehouses. In cases involving unbounded or very large amount of data, the original detailed data are replaced by aggregated ones, at a granularity level that is a trade-off between analytical needs for pinpointing situations discovered at macro level and system resources. In [28] Han et al. extend the concept of data warehousing to spatial data (SDW) and introduce the concept of spatial data cube. Their goal is to balance the storage requirements and cost of aggregation at query-time by precomputing part of aggregates. Also in [40] the focus is on pre-aggregation in SDW. Both [28] and [40] present methods suitable only for aggregating facts whose measures are distributive over the aggregation operator. In [34] a conceptual multidimensional data model for spatial data is presented. It allows for spatial dimensions, spatial hierarchies, and spatial measures, as well as its logical representation based on a commercial spatial DBMS [35]. In [53] Spatial OLAP operators are used to analyse the acceleration and speed of skaters along a race track. In [21] a framework, named Piet, is proposed for the integration of GIS and OLAP. In [8] the usual Spatial OLAP operators are extended with the addition of five new ones that allow to navigate also the hierarchy of spatial measures and to modify the structure of the geographic hypercube during the analysis process, as opposed to the usual a priori definition of the dimensional hierarchies. The work [47] addresses the issue of ad hoc spatial query performances in SDW, by exploiting domain indexes specifically designed for spatial data warehouses.

The evolution of SDW naturally leads to include in the model a temporal dimension for dynamic spatial data. In [29] a conceptual model for moving objects with imprecise position is presented, and it can be instantiated with different dimensions, hierarchies and measures. In [37] we described a data model for storing measures related to trajectories, focusing on the efficient approximation of aggregates. We adopted the same data model in [36], to evaluate design solutions that integrate moving object databases (MOD) and TDW, and in [44] we used the proposed framework (MOD and TDW) to examine traffic data, in combination with tools for the visual analysis of spatio-temporal data. The main limitation of the mentioned approaches is the fact that they are restricted to freely moving objects. Thus, they do not allow to explicitly account for constrained movements, for example due to the presence of a road network. As illustrated in the sequel (see Section 7) this can seriously compromise the quality of the analysis. [52] presents a multidimensional model for the representation of data related to location based services for vehicles moving along a road network. As the authors observe, their work is tailored to the requirements of Location-Based Services (LBSs) for objects moving on a network. In [54] an OLAP system for network-constrained moving objects is introduced. The proposed system is based on the efficient indexing of individual trajectories, and thus

it is able to answer detailed queries. On the other hand, this approach hinders the use of the system for very large databases. In [33] a method to efficiently process large scale, real-time, traffic data and update aggregate summaries related to road segments is presented. We observe, however, that the contribution of this work is mainly focused on the aggregation of raw data to detect the road segments presenting a significant variation of their average speed. Indeed, their approach is orthogonal to the one we propose, and could be adopted to feed our TDW. Finally, in [22], Gómez et al. present an extension of their Piet framework to deal with trajectory data.

In the analysis of spatial and spatio-temporal data, the use of suitable, interactive, visualization tools is of paramount importance to help the analytic user in effectively grasping the information hidden in those complex data. In this context, data aggregation is commonly used for dealing with large amounts of data. In particular, spatial, temporal, and categorical aggregations are used for spatio-temporal data [18]. A survey of the aggregation methods used for movement data is proposed in [3]. To study the distribution of movement characteristics over space, movement data are aggregated into continuous density surfaces (e.g. [13]) or discrete grids (e.g. [3, 17]). Brillinger et al. in [10] aggregate movement data into a vector field using a regular grid: in each grid cell, a vector is drawn with an angle corresponding to the prevailing movement direction and length and width proportional to the average speed and the amount of movement, respectively. To study links between places, movement data are aggregated into origin-destination matrices [24] and flow maps [5].

Visual OLAP is a clear trend in software for business visualization. The tools allow the user to explore data cubes through traditional visualisation techniques such as time series plots, scatterplots, maps, treemaps, cartograms, matrices etc., as well as more specialised visualisations. Polaris and ADVIZOR are two pioneering systems in this direction. Polaris [48] is a visual tool for multidimensional analysis developed at Stanford University. Currently, Tableau Software commercialises the Polaris work. ADVIZOR represents the commercialisation of 10 years of research in Bell Labs on interactive data visualisation and in-memory data management [15]. To our knowledge, there are no specific visual analytics systems working with spatio-temporal data warehouses.

Visualization of large amounts of spatio-temporal data is a challenging research topic in visual analytics, as pointed out in the recently published roadmap for the visual analytics research [31]. To address this problem, visual analytics requires support from the data management side: architectures for data management, specialized query languages and operators, dynamic processing and efficient algorithms to keep an up-to-date online connection to the data sources. It is also necessary to design efficient algorithms for analysis of dynamic (streaming) data, in particular, algorithms that are able to proceed in an incremental way and capture both trends and overall insights. An overview of the state of the art in visual analytics of movement data is provided in [7], including methods for analysing trajectories (visualization, clustering, interactive transformation of temporal references), studying trajectory attributes and event extraction from trajectory data, generalization and aggregation of movement, and investigation of movement in context.

3 Application scenarios

Before defining the model of our TDW and the developed visual analysis tools, we introduce some significant application scenarios, with the goal of highlighting issues

and needs that may arise in different contexts. The first scenario is about freely moving objects, namely ships sailing on the sea. The second one deals with objects whose movement is constrained to a network, e.g., cars moving along roads.

3.1 Vessels sailing on the sea

In this scenario we are interested in analysing the movements of vessels which can be considered as an example of unconstrained free motion. Indeed these movements are not completely unconstrained, due to land presence, reduced water depth, or traffic constraints, however, ships can move freely in the open sea.

As a data source we used the Vessel Monitoring System (VMS) which has been established since 2005 by the European Union. The legislation requires that all fishing vessels transmit vessel identification, date, time, position, course and speed hourly. VMS is mainly intended to monitor the movement of vessels with respect to restricted fishing areas in real time.

In order to reconstruct the trajectories of vessels starting from VMS data, i.e., to derive a global function of time describing the whole trajectory, *local interpolation* can be used. According to this method, objects are assumed to move between the observed points following some rule. For instance, a *linear* interpolation function models a straight movement with constant speed, while other polynomial interpolations can represent smooth changes of direction.

The scenario has been concretely studied using a private dataset containing the positions of ships sailing on the North Adriatic Sea between January and September 2007. It consists of 326,800 records concerning the movements of 270 boats (i.e. about 63,000 trajectories). Moreover, we used an additional dataset storing information about the amount of fish caught by each boat and sold on the market every day. This dataset contains around 141,000 records. Each record includes the date of the sale, the fish species, the boat which captured the fishes, and the amount of fish sold (in kilos). These data are used in order to estimate the catches of each boat on its daily path. Catches are then assumed to be uniformly distributed along the portions of the path classified as "fishing" (the classification is based on several parameters including the distance from the coast and the boat speed). The analyses of interest to be performed on these data have been identified in collaboration with a group of environmental scientists. Their main interests are concerned with the *fishing effort index*, i.e., a value indicating how much a given area has been exploited by the boats fishing in it, and the *distribution of the species* on the sea. Some typical queries are: *"How are species distributed on the sea?" "Which are the most exploited zones?" "Which are the zones with the highest amount of catches/fishing effort ratio?"*.

3.2 Cars moving along a road network

In this scenario we are interested in cars moving along a road network, which is modelled as a graph embedded in the Euclidean 2D-space. Positions are collected by on board GPS devices at irregular rates. In this case the movement is completely constrained since cars are supposed to stay on the network. When reconstructing their trajectories, we should take into account the topology of the road network to determine the path followed by each car between two consecutive GPS positions (see e.g., [9]). The reconstruction phase produces a sequence of lines in a $3D$ space

($\mathbb{T} \times \mathbb{R}^2$), each representing the continuous "development" of the moving object during a time interval. Notice that the spatial projection of these lines are road segments of the road network or portions of these segments.

A typical user of this scenario is a city manager who exploits the TDW to analyse the traffic. Some possible queries are: *"Which is the number of buses per hour in the morning of a given day in the neighbourhoods of a given district? Show its temporal evolution using a temporal granularity of half an hour"*, or *"From which district does a great number of cars leave in the morning? And at what hour? Is there a flow exiting/entering the town? Which are the main differences in the traffic between the working days and the week-end?"*.

As a concrete instance of this scenario we analyse a large real world dataset which contains the observations of GPS-equipped cars moving in the urban area of Milan (Italy). This is a private dataset used in the GeoPKDD EU Project [19]. It consists of two millions of records that represent the movement of 17,000 objects (i.e. about 200,000 trajectories) moving during a week period from Sunday to Saturday.

4 TDW conceptual model

Our TDW has to model facts related to multitudes of moving object trajectories. Since we are interested in reporting aggregate measures at different levels of granularity, which refer to specific spatial zones during given temporal periods, the dimensions of analysis have to include both space and time.

A first choice regards the TDW *base granularity*, which is a collection of elements called *base granules*, obtained by partitioning both the spatial and temporal dimensions. Informally, a granule can be defined as a contiguous spatial region during a given time interval. According to the different scenarios, the granule-based decomposition can generate a regular or irregular tessellation of the space domain during a given time interval. For example, in the vessel scenario, the North Adriatic Sea can be partitioned in squared regular areas whereas in the road traffic scenario, base granules can be associated with segments of the road network.

Further dimensions of analysis can be added to the TDW in order to group trajectories based on various features of the moving objects. Also in this case, the dimensions can differ depending on the application scenarios, e.g., they can include demographic information if we are considering people's trajectories, or vessel typology if we are dealing with ship trajectories.

The TDW will then store aggregate measures concerning quantitative aspects of sets of trajectories crossing the spatio-temporal base granules and belonging to specific object groups. These sub-aggregates at base granularities can be further aggregated to answer roll-up queries.

Conceptual model In Fig. 1 we present a conceptual model for a TDW, built by using the Dimensional Fact Model formalism [20]. Facts, focus of interest of the decision process, are represented by boxes containing the fact name and a list of associated measures (quantitative aspects interesting for analysis). In our model we consider two classes of facts, namely INTRA-GRANULE and INTER-GRANULE facts. In the following we introduce such facts and a set of significant measures, which will be formally defined in the next subsections.

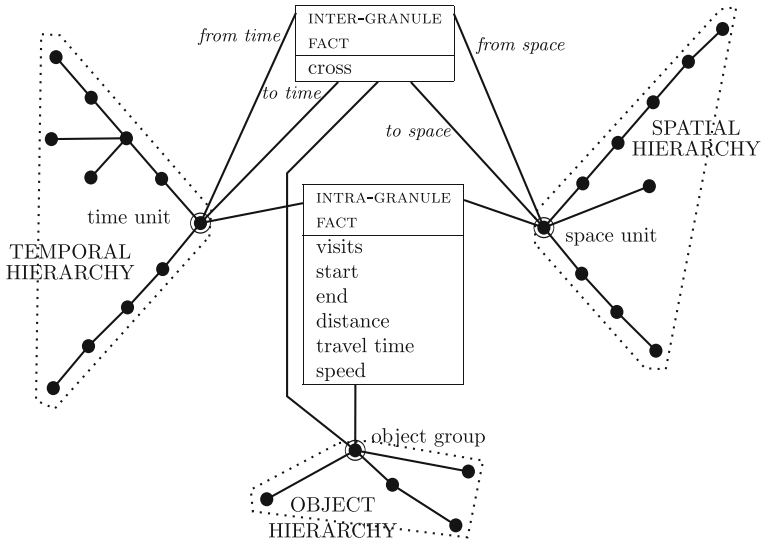


Fig. 1 TDW Conceptual model

INTRA-GRANULE: These facts model events that are related to a *single base granule* concerning a certain object group. For a given group U and a granule g, the measures are:

- *visits*: the number of trajectories belonging to group U which start from or enter into granule g;
- *start/end*: the number of trajectories belonging to group U starting/ending in granule g;
- *travel time/distance*: the time spent/distance travelled by all trajectories belonging to group U while moving inside granule g;
- *speed*: the average speed of trajectories belonging to group U traversing granule g.

INTER-GRANULE: These facts model events that are related to *pairs of granules* and are concerned with a specific object group. For a given group U and pair of granules g and g', a measure of interest is

- *cross*: number of times the border from g to g' has been traversed by trajectories belonging to a group U.

Note that the measure *cross* is interesting only for adjacent granules (for non-adjacent granules it is invariably 0). However, in general, inter-granule facts can model events which are meaningful for all pairs of granules. An example could be the *origin-destination* measure, which, for any pair of granules, represents the number of trajectories starting from the first and ending into the second granule.

Clearly, the presented measures are not an exhaustive collection, but they correspond to a set of common measures which we found interesting and useful in different scenarios.

Dimensions, i.e., the coordinates of analysis of facts at the finest level of granularity, are represented as circles attached to the fact tables by straight lines. As already mentioned, the dimensions in our model are a *spatial* and a *temporal* dimensions describing geography and time, respectively, and a non spatio-temporal dimension (*object group*). The choice of considering a single dimension for modelling space, time and object group is only aimed at keeping the notation simple: there would be no conceptual obstacle in considering multiple dimensions. Indeed, this normally happens in concrete instantiations of the framework. For instance, in Section 7.1 for the vessel case study, the object group is modelled by using three distinct dimensions (boats, tools and activities).

Dimensional attributes, i.e. the properties of dimensions, are represented as circles attached directly to dimensions or to other dimensional attributes. A dimensional hierarchy consists of a dimension and its dimensional attributes, hence represented as a rooted tree, having the root attached to a fact table.

For the sake of readability, hierarchies that are common to various dimensions are shared in the graphical representation. This is indicated using a double circle for the root node. Moreover, when a dimension appears more than once in a fact table, it is necessary to specify a role to define its meaning. For example, in Fig. 1 *from space* and *to space* associated with the spatial dimension of the fact table Inter-Granule denote, respectively, the origin/destination granule of the movement.

4.1 Examples of spatio-temporal hierarchies

An example of a complex temporal hierarchy is illustrated in Fig. 2. We consider *minutes* as the minimal temporal interval. Every minute belongs to a 10 *minutes interval*. In turn, each 10 *minutes interval* is contained in an *hour*, which is included in one *day*. A day is contained in both a *week* and a *month*, and it is also a *day of the week*. Moreover, the 10 *minutes interval* is included also in intervals of larger and larger duration.

The spatial hierarchy, in the case of the vessel scenario, may be based on a collection of regular grids of increasing size, such as the one represented in Fig. 3a, whereas for the road traffic scenario we can have a hierarchy like the one represented in Fig. 3b, where the segment is the smallest spatial unit. Each segment is contained in exactly one *district*, which is included in one *zone* belonging to a *province*, and in turn contained in a *region* and finally in a *country*. At the same time, each segment is included into a *cell* of a 200 m $\times$ 200 m grid, which is contained in a *cell* of a

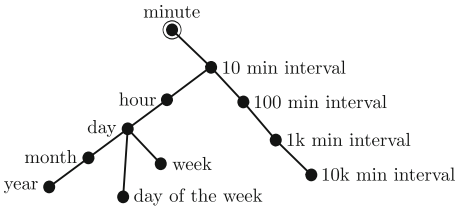


Fig. 2 An example of a temporal hierarchy

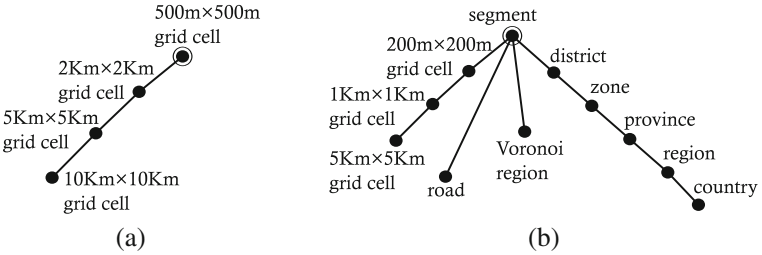


Fig. 3 Examples of spatial hierarchies for the vessel and car scenarios

1 km × 1 km grid, and so on. Finally a segment is also inside a Voronoi region built by using the distribution of the raw trajectory data.

5 TDW spatio-temporal hierarchies and measures

In this section we first define a discretization of space and time domains of our TDW into so-called *spatio-temporal granules*, and then their organization in *hierarchies*. We also introduce the concept of *trajectory decomposition* according to the discretized spatio-temporal domain. We will exploit this concept in order to provide a rigorous definition of the TDW *measures* informally presented in the previous section. Finally, we define the aggregate functions which allow us to implement roll-up operations in the TDW.

5.1 Granules, granularities and hierarchies

We need to discretize the space and time dimensions in order to define the TDW base granularity. We start discussing the discretization of a generic domain.

Definition 1 (Granule and Granularity) Let $\mathbb{D}$ be any domain (e.g. temporal or spatial). A *granularity* G on $\mathbb{D}$ is any *partition* of $\mathbb{D} = \bigcup_{g \in G} g$, whose elements g are called *granules*.

Given two granularities G and G' we say that G is *finer* than G' and write $G \preceq G'$, if for all $g \in G$, there exists $g' \in G'$ such that $g \subseteq g'$. In this case we also say that G' is *coarser* than G.

Definition 2 (Hierarchy) A *hierarchy* over a domain $\mathbb{D}$ is a set $\mathscr{H}$ of granularities on $\mathbb{D}$, partially ordered by $\preceq$, and with a minimum, called the *base granularity*, composed of *base granules*.

In the following we assume that for the spatial domain $\mathbb{S}$ a finite hierarchy is fixed, denoted by $\mathscr{H}_S$, with base granularity $G_{S\perp}$, and such that for any $G_S \in \mathscr{H}_S$ and each

granule $g_S \in G_S$, g_S is topologically connected. Similarly, for the temporal domain $\mathbb{T}$ we fix a finite hierarchy, denoted by $\mathscr{H}_T$, with base granularity $G_{T\perp}$, and such that each granule at any granularity is a temporal interval. Observe that these induce a finite hierarchy over the spatio-temporal domain $\mathbb{T} \times \mathbb{S}$, resulting as $\mathscr{H}_{TS} = \{G_T \times G_S \mid G_T \in \mathscr{H}_T \ \wedge \ G_S \in \mathscr{H}_S\}$. Hereinafter we denote a granule in $\mathbb{T} \times \mathbb{S}$ as a pair (g_T, g_S). In addition, when taking a granularity $G \in \mathscr{H}_{TS}$, we will denote by G_T and G_S the corresponding temporal and spatial granularities, such that $G = G_T \times G_S$. Moreover we will write $G_\perp$ for the base granularity $G_{T\perp} \times G_{S\perp}$.

5.2 Trajectory decomposition

In the following we will refer to sets of trajectories indexed by trajectory identifiers $\mathsf{T} = \{\mathsf{T}_{id}\}_{id \in \mathscr{I}}$ where $\mathscr{I}$ is the set of trajectory identifiers. For any trajectory identifier $id \in \mathscr{I}$, the corresponding trajectory is a function $\mathsf{T}_{id} : I_{id} \to \mathbb{R}^2$, where I_{id} is the time interval of definition of the trajectory. Equivalently, T_{id} can be seen as an infinite set of points in a 3D space $\{(t, \mathsf{T}_{id}(t)) \mid t \in I_{id}\}$, i.e., the graph of the function. In the sequel we will use these two views interchangeably.

As mentioned before, we also assume the presence of object groups, where a generic group will be denoted as U and, abusing the notation, we will write $id \in U$ to mean that the (object corresponding to trajectory) id belongs to group U.

Definition 3 (trajectory decomposition) Let T be a set of trajectories and let $G_\perp$ be a base granularity. For each trajectory $\mathsf{T}_{id} : I_{id} \to \mathbb{R}^2$ in T the *trajectory decomposition* is a sequence of sub-trajectories $\delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle$, with $s^i_{id} : I^i_{id} \to \mathbb{R}^2$ and $I^i_{id} = [t_{start}{}^i_{id}, t_{end}{}^i_{id})$, satisfying the following conditions:

- $\sup(I^i_{id}) = \inf(I^{i+1}_{id})$ for all $i \in \{1, \ldots, n-1\}$,
- $\bigcup_{i=1}^n s^i_{id} = \mathsf{T}_{id}$,
- for any $i \in \{1, \ldots, n\}$, there exists a granule $g \in G_\perp$ such that $s^i_{id} \subseteq g$.

Informally, $\delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle$ is a partition of the trajectory T_{id}, where each s^i_{id} is a fragment of the trajectory included in a granule of $G_\perp$, such that each s^i_{id} temporally precedes s^{i+1}_{id}.

Example 1 Let us consider the road traffic scenario as a running example. The spatial domain is a road network RN embedded in $\mathbb{R}^2$, i.e. $RN \subseteq \mathbb{R}^2$.

Figure 4 shows a part of a user trajectory T_{id}, depicted as a solid polyline in the centre of the route. The arrows indicate the direction of the movement. Notice that the movement of the user is constrained to the roads (in white), and grey areas are unreachable.

We take as base spatial granules $\{g_{e1}, \ldots, g_{e8}\}$, corresponding to road segments of the road network, while the base temporal granularity consists of regular intervals of 20 time units starting from 0 ($\{[0, 20), [20, 40), \ldots\}$). Hence according to this spatio-temporal granularity the decomposition of T_{id} is

$$\delta(\mathsf{T}_{id}) = \langle s_1, \ldots, s_9 \rangle$$

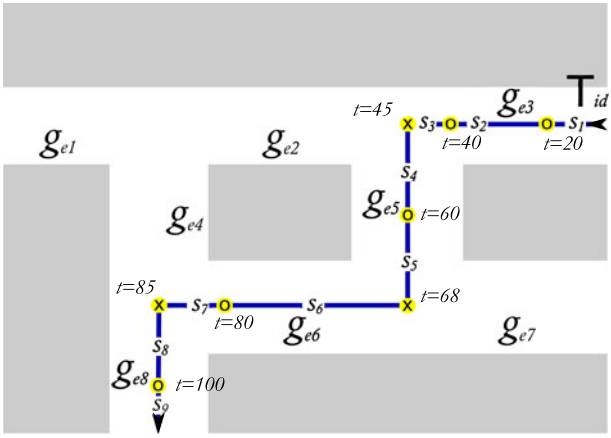


Fig. 4 Decomposition of a trajectory

The symbols ○ and × indicate the points $(t, \mathsf{T}_{id}(t))$ where the trajectory enters, respectively, a new temporal granule or a new spatial granule, and the label indicates the corresponding time t.

5.3 Intra-granule measures

A first intra-granule measure is *visits* ($\mathscr{V}$), which represents the number of visits in g of the trajectories belonging to a certain group U. More precisely, $\mathscr{V}^{\mathsf{T}}(g, U)$ is defined as the number of times that any trajectory belonging to U enters into, or starts from the spatio-temporal granule g.

Definition 4 (Visits) Let G be a spatio-temporal granularity and $g \in G$ be a granule, let T be a set of trajectories and let U be an object group, then

$$\mathscr{V}^{\mathsf{T}}(g, U) = \left|\left\{\left(id, s^i_{id}\right) \mid id \in U \ \wedge \ \delta\left(\mathsf{T}_{id}\right) = \langle s^1_{id}, \ldots, s^n_{id}\rangle \ \wedge \ s^i_{id} \subseteq g \ \wedge \ \left(i = 1 \ \vee \ s^{i-1}_{id} \not\subseteq g\right)\right\}\right|$$

Informally, a trajectory visits a granule $g = (g_T, g_S)$ if there exists $t \in g_T$ such that the trajectory is in g_S at time t, but it was not in g_S immediately before t, either because the trajectory starts at time t, or the trajectory was in another spatial granule. Moreover, a trajectory can visit g even without any movement: an object, that stays in g_S for a long time period, can enter different spatio-temporal granules having the same spatial granule g_S but different temporal granules g_T.

Other intra-granule measures are *start* ($\mathscr{S}$) and *end* ($\mathscr{E}$), i.e., the number of trajectories of a given group starting and ending in a granule, *distance* (*dist*), *travel time* (*trav_t*) and *speed*, i.e., the total travelled distance by the trajectories of a given group, the total time spent in the granule and the corresponding average speed. Their definitions, which are given below, are the natural ones.

Definition 5 (Other measures) Let G be a spatio-temporal granularity and $g \in G$ be a granule, let T be a set of trajectories and let U be an object group, then

$$\mathscr{S}^{\mathsf{T}}(g, U) = \left|\left\{id \mid id \in U \wedge \delta(\mathsf{T}_{id}) = \left\langle s^1_{id}, \ldots, s^n_{id}\right\rangle \wedge s^1_{id} \subseteq g\right\}\right|$$

$$\mathscr{E}^{\mathsf{T}}(g, U) = \left|\left\{id \mid id \in U \wedge \delta(\mathsf{T}_{id}) = \left\langle s^1_{id}, \ldots, s^n_{id}\right\rangle \wedge s^n_{id} \subseteq g\right\}\right|$$

$$dist^{\mathsf{T}}(g, U) = \Sigma_{id \in U,\, \delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id}\rangle}\, \Sigma_{s^j_{id} \subseteq g} len\left(s^j_{id}\right)$$

$$trav\_t^{\mathsf{T}}(g, U) = \Sigma_{id \in U,\, \delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id}\rangle}\, \Sigma_{s^j_{id} \subseteq g} lifespan\left(s^j_{id}\right)$$

$$speed^{\mathsf{T}}(g, U) = \frac{dist^{\mathsf{T}}(g, U)}{trav\_t^{\mathsf{T}}(g, U)}$$

where $len(s^j_{id})$ is the length of the spatial projection of s^j_{id} whereas $lifespan(s^j_{id})$ is the duration of the time interval where s^j_{id} is defined.

5.4 Inter-granule measures

The only inter-granule measure we discuss in this paper is *cross* ($\mathscr{C}$), the number of times the border from a granule to another has been traversed by trajectories of a given group.

Definition 6 (Cross) Let G be a spatio-temporal granularity, let $g, g' \in G$ be two distinct granules, let T be a set of trajectories and let U be an object group, then

$$\mathscr{C}^{\mathsf{T}}(g, g', U) = \left|\left\{(id, s^i_{id}) \mid id \in U \wedge \delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id}\rangle \wedge i < n \wedge s^i_{id} \subseteq g \wedge s^{i+1}_{id} \subseteq g'\right\}\right|$$

The measure $\mathscr{C}$ is not symmetric: in general $\mathscr{C}^{\mathsf{T}}(g, g', U) \neq \mathscr{C}^{\mathsf{T}}(g', g, U)$ since the measure is sensible to the direction of movements and only counts crossings from g to g'. Note that $\mathscr{C}^{\mathsf{T}}(g, g', U) = 0$ when g and g' are not *adjacent*, where the *adjacency* relation is defined in the usual way. Let $g = (g_T, g_S)$ and $g' = (g'_T, g'_S)$ two granules belonging to a granularity G; they are *adjacent* when $g_T = g'_T \wedge Touch(g_S, g'_S)$ or $g_S = g'_S \wedge Meets(g_T, g'_T)$ or $Meets(g_T, g'_T) \wedge Touch(g_S, g'_S)$ where *Meets* is Allen's relation [1] and *Touch* is Egenhofer's topological relation [14].

Example 2 We still refer to the running example of a TDW for road traffic analysis. Figure 5a illustrates portions of two trajectories, T_{id1} and T_{id2}, belonging to the same group U, during a temporal granule g_T. The direction of a trajectory is indicated by an arrow. The trajectory T_{id1}, in light green (light gray in B&W), passes through the sequence of base spatial granules g_{e1}, g_{e4}, g_{e6}, g_{e5}, g_{e2}, g_{e4}, g_{e8}, whereas T_{id2}, in blue (black in B&W), travels along g_{e3}, g_{e5}, g_{e6}, g_{e8}.

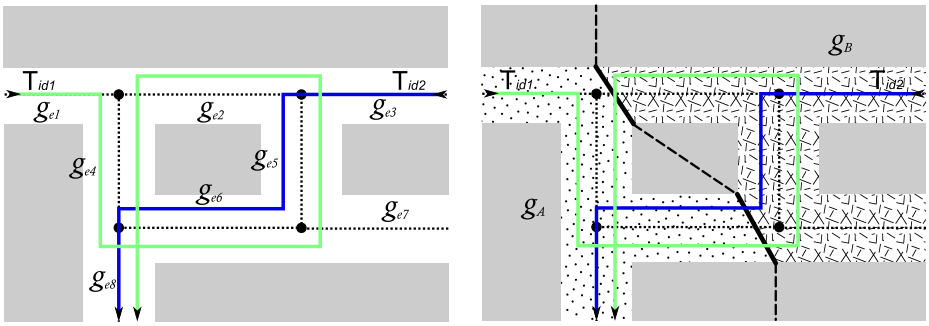


Fig. 5 Two trajectories at base **a** and higher **b** spatial granularity

Note that $\mathscr{V}^{\mathsf{T}}((g_T, g_{e4}), U) = 2$ since T_{id1} enters twice into g_{e4}, while T_{id2} does not enter into such a granule. Instead, $\mathscr{V}^{\mathsf{T}}((g_T, g_{e6}), U) = 2$ since both trajectories enter only once into g_{e6}.

Focusing on the crosses between the granule (g_T, g_{e2}) and the adjacent granules, we have:

$$\mathscr{C}^{\mathsf{T}}((g_T, g_{e2}), (g_T, g_{e4}), U) = \mathscr{C}^{\mathsf{T}}((g_T, g_{e5}), (g_T, g_{e2}), U) = 1$$

while the remaining crosses are all equal to 0.

5.5 Aggregation

In order to allow for OLAP processing, our TDW has to offer aggregation capabilities over measures, i.e., operations for computing measures at some higher level of the hierarchy starting from those at lower level. Efficient OLAP roll-up operations require that measures at a coarser granularity can be determined using values at a finer granularity. For every measure m, we thus provide an inductive characterization of such roll-up operation (proofs in the Appendix A).

Proposition 1 *Let $\mathscr{H}_{TS}$ be a hierarchy, let T be a set of trajectories and U be an object group. For any $G \in \mathscr{H}_{TS}$ with $G \neq G_\perp$, $g, g' \in G$ with $g \neq g'$*

$$\mathscr{V}^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} \left(\mathscr{V}^{\mathsf{T}}(g_p, U) - \Sigma_{g'_p \subseteq g} \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) \right)$$

$$\mathscr{S}^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} \mathscr{S}^{\mathsf{T}}(g_p, U)$$

$$\mathscr{E}^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} \mathscr{E}^{\mathsf{T}}(g_p, U)$$

$$\mathscr{C}^{\mathsf{T}}(g, g', U) = \Sigma_{g_p \subseteq g, g'_p \subseteq g'} \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U)$$

$$dist^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} dist^{\mathsf{T}}(g_p, U)$$

$$trav\_t^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} trav\_t^{\mathsf{T}}(g_p, U)$$

where $g_p, g'_p \in G_p$ with $g_p \neq g'_p$ and G_P is a predecessor of G, i.e., $G_p \preceq G$ and $G_p \neq G$.

The inductive characterisation above is self-explanatory for most measures. Some intuitive explanation is needed for the aggregate function for measure $\mathscr{V}$. At a coarser granule g the number of visits in g is obtained by summing up the visits in the finer granules g_p composing g, and subtracting the number of trajectories crossing the border between two distinct finer granules inside g. This is motivated by the fact that the border between two finer granules, g_p and g'_p composing g is completely inside g. Hence trajectories moving from g_p to g'_p (or vice versa) increase the number of visits in g'_p (or g_p) but they should not be counted as visits in the coarser granule g because the movement is completely inside g, i.e., they do not *enter* g.

Example 3 Figure 5b focuses on the same temporal granule g_T of Fig. 5a. The spatial granularity is made coarser by considering two granules $g_A = g_{e1} \cup g_{e4} \cup g_{e6} \cup g_{e8}$ and $g_B = g_{e2} \cup g_{e3} \cup g_{e5} \cup g_{e7}$

Trajectory T_{id1} starts from the granule g_A, traverses the granule g_B, and finally ends inside the granule g_A, whereas T_{id2} begins in the granule g_B and ends in the granule g_A. Thus, $\mathscr{V}^{\mathsf{T}}((g_T, g_A), U) = 3$, $\mathscr{V}^{\mathsf{T}}((g_T, g_B), U) = 2$, $\mathscr{C}^{\mathsf{T}}((g_T, g_A), (g_T, g_B), U) = 1$, and $\mathscr{C}^{\mathsf{T}}((g_T, g_B), (g_T, g_A), U) = 2$.

Now we want to apply Proposition 1 to compute $\mathscr{V}^{\mathsf{T}}((g_T, g_A), U)$. The granules composing g_A are $g_{e1}, g_{e4}, g_{e6}, g_{e8}$ and $\mathscr{V}^{\mathsf{T}}((g_T, g_{e1}), U) = 1$, $\mathscr{V}^{\mathsf{T}}((g_T, g_{e4}), U) = \mathscr{V}^{\mathsf{T}}((g_T, g_{e6}), U) = \mathscr{V}^{\mathsf{T}}((g_T, g_{e8}), U) = 2$. The only non-null crosses between granules contained in g_A are $\mathscr{C}^{\mathsf{T}}((g_T, g_{e1}), (g_T, g_{e4}), U) = \mathscr{C}^{\mathsf{T}}((g_T, g_{e4}), (g_T, g_{e6}), U) =^{\mathsf{T}} ((g_T, g_{e4}), (g_T, g_{e8}), U) = \mathscr{C}^{\mathsf{T}}((g_T, g_{e6}), (g_T, g_{e8}), U) = 1$. Hence by adding the $\mathscr{V}$'s and subtracting the $\mathscr{C}$'s we obtain $7 - 4 = 3$, which is the exact value for $\mathscr{V}^{\mathsf{T}}((g_T, g_A), U)$.

The above proposition suggests the aggregate functions which can be used to compute a measure m at coarser granularities by exploiting the sub-aggregates for finer granularities. In the fact tables only facts with non-empty values are stored, so e.g., the Inter-Granule Fact only includes *adjacent* granules with non-empty crosses of the borders. Moreover, the spatial and temporal hierarchies are fixed, hence for each base granule the temporal and spatial coarser granules it belongs to can be pre-computed. In this way, the containment condition, i.e., establishing whether a granule belongs to a coarser one, does not require any spatio-temporal query but only an equality test on a constant value.

According to the classification by Gray et al. [23], the aggregate functions for $\mathscr{S}$, $\mathscr{E}$, $\mathscr{C}$, *dist* and *trav_t* are *distributive*, i.e. super-aggregates are computed by summing up the sub-aggregates at finer granularities. On the other hand, the aggregate function for $\mathscr{V}$ is *algebraic* because super-aggregates are computed from the sub-aggregates with a *finite* set of auxiliary measures. The same applies to *speed* which is computed by using the auxiliary measures *dist* and *trav_t*.

6 Approximating presence

In this section we discuss the use of measure $\mathscr{V}$ to estimate *Presence*, i.e. the number of *distinct* trajectories belonging to a certain object group travelling in a given spatio-temporal granule.

We start by providing a formal definition of measure *Presence* ($\mathscr{P}$).

Definition 7 (Presence) Let G be a spatio-temporal granularity and $g \in G$ be a granule, let T be a set of trajectories and U be an object group, then

$$\mathscr{P}^{\mathsf{T}}(g, U) = \left|\left\{id \in U \mid \delta(\mathsf{T}_{id}) = \left\langle s^1_{id}, \ldots, s^n_{id}\right\rangle \wedge \exists i \in \{1, \ldots, n\}. s^i_{id} \subseteq g\right\}\right|$$

This measure is hard to handle in a TDW since the corresponding aggregate function is *holistic*: the raw data are needed to compute the exact result at all granularities. This is due to the fact that trajectories might span multiple granules. Hence in the aggregation phase we have to cope with the so called *distinct count problem* [50]: if an object remains in the query region for several timestamps during the query interval, one should avoid to count it multiple times in the result. This is problematic since, once loaded in the TDW, the identifiers of the trajectories are lost.

Several approaches have been proposed in the literature to provide an approximation of measure $\mathscr{P}$, using aggregated data. For instance, a distributive function that simply sums the *Presence* at the finer levels to compute the super-aggregate at a coarser granularity has been considered (e.g., [38]). A more refined approach based on so-called FM-sketches[1] is discussed in [50]. It relies on a probabilistic counting that makes use of bit vectors. The number of bit vectors affects the accuracy.

Interestingly enough, as mentioned above, the measure $\mathscr{V}$, computed as in Section 5, can be profitably used for approximating $\mathscr{P}$. A simple but key observation is that $\mathscr{P}$ differs from $\mathscr{V}$ for the fact that multiple visits of the same trajectory to granule g are counted once.

Example 4 Consider again Fig. 5a. We have that $\mathscr{P}^{\mathsf{T}}((g_T, g_{e4}), U) = 1$ whereas $\mathscr{V}^{\mathsf{T}}((g_T, g_{e4}), U) = 2$.

From the definitions of $\mathscr{V}$ and $\mathscr{P}$ it is immediate to see that $\mathscr{V}$ is an upper bound for $\mathscr{P}$.

Proposition 2 *Let G be a spatio-temporal granularity and $g \in G$ be a granule, let T be a set of trajectories and U be an object group, then $\mathscr{P}^{\mathsf{T}}(g, U) \leq \mathscr{V}^{\mathsf{T}}(g, U)$.*

It is worth noting that the difference between $\mathscr{V}$ an $\mathscr{P}$ is larger when trajectories are very *agile*, i.e., they frequently change their direction. This happens because the same trajectory can get back to a granule that it already visited. The phenomenon disappears when we increase the size of granules by rolling-up, since at some point all trajectories will be completely contained in a granule. This intuition is formalised in the following proposition (a more formal statement and its proof can be found in Appendix A).

Proposition 3 *Let G be a spatio-temporal granularity and $g \in G$, let T be a set of trajectories and let U be a user group, then*

1. *if each trajectory visits g at most once then $\mathscr{P}^{\mathsf{T}}(g, U) = \mathscr{V}^{\mathsf{T}}(g, U)$;*
2. *if all the trajectories are inside g then $\mathscr{P}^{\mathsf{T}}(g, U) = \mathscr{V}^{\mathsf{T}}(g, U)$.*

[1] FM from the Flajolet and Martin, the original proposers' names

According to the first statement if any trajectory visits a granule g at most once then $\mathscr{P}$ coincides with $\mathscr{V}$. In particular, if $\mathscr{V}^{\mathsf{T}}(g, U) = 0$ then $\mathscr{P}^{\mathsf{T}}(g, U) = 0$. The second statement suggests that the same happens for coarse granularities.

Experimentally, it can be seen that $\mathscr{V}$ provides a very accurate estimate of $\mathscr{P}$, which outperforms the other approaches mentioned above, i.e., the distributive function adding the values at coarser granularities (e.g., [38]) and sketches [50].

The experimental comparison reported in Fig. 6 has been obtained by using the dataset described in Section 3.2. As base granularity, we set a grid of rectangles, of size 330 m $\times$ 440 m, and time intervals of 1 hour. In order to compare the errors we chose to adopt as an aggregation accuracy metric the normalised absolute error defined as follows:

$$Error = \frac{\sum_g Error(g)}{\sum_g g.Pres} = \frac{\sum_g |\widetilde{g.Pres} - g.Pres|}{\sum_g g.Pres} \tag{1}$$

where g are granules at a coarser granularity than the base one, $g.Pres$ is the exact value of *Presence* in the granule g whereas $\widetilde{g.Pres}$ is the approximated value obtained using one of the discussed methods, i.e. distributive function, FM sketches and the measure $\mathscr{V}$.

The graphs in Fig. 6 show the normalised absolute errors as functions of the granularities. The values indicated for the granularities are relative to the base one. For example, a value 2 for granularity means that we are considering granules having double size w.r.t. the base granules along all dimensions. The different curves for FM sketches correspond to a different number, m, of bit vectors.

As shown by the corresponding curves, the *distributive* aggregate function (the top curve) quickly reaches very large errors as the roll-up granularity increases. This is due to the fact that we simply sum the sub-aggregates and as a consequence trajectories crossing different granules are counted many times: the number of duplicates becomes higher and higher at coarser granularities. Interestingly, for all

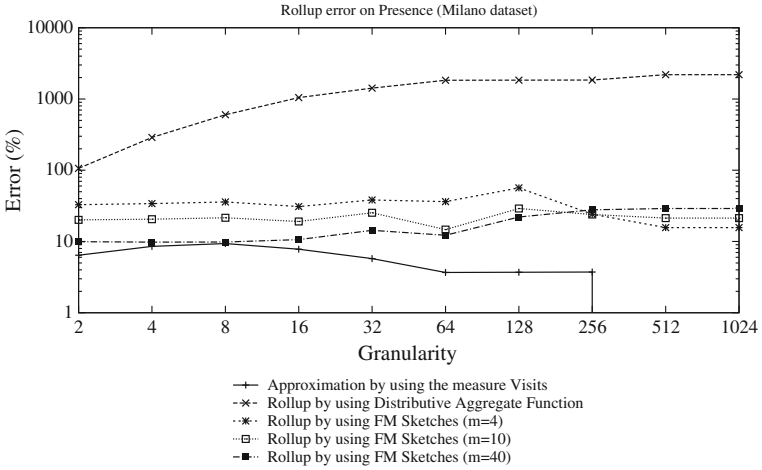


Fig. 6 Cumulative error of roll-up phase

granularities $\mathscr{V}$ outperforms also FM sketches and, for coarser ones, $\mathscr{V}$ is no longer an approximation but it coincides with the measure *Presence*.

7 Visual OLAP

Multidimensional analysis is a typical example of analytics that can be supported by a TDW. Some typical questions taken from our scenarios are: "When and in which area/roads of the town does the most intense traffic appear?", "Which are the zones with the highest number of ships entering from different directions per hour?", "Is there any difference in traffic between the working days and the week-end?", "How does the movement propagate from place to place?". Standard, table based OLAP operations could be used to answer some of these queries, some others require TDW for handling the specifics of movement. However, the interpretation of results, and the consequent refinement of queries and exploration of results, is not easy. Integrating OLAP tools with visualization provides advanced analysis capabilities. For instance, trajectory data can be geo-referenced in a map and combined with several layers (such as topography, demography, other themes). Performing OLAP operations on TDW specialised measures and using visual tools make the exploration of the data cube more rapid and intuitive. For these reasons, we have provided the TDW with an interface that allows for OLAP visual operations, based on V-Analytics [4, 7], an interactive Java™ based visual analytics system. This system permits a user to view georeferenced data over a map and run analyses on them, for example to find clusters or to tessellate the space. It also offers functionalities to handle temporal data, by using graphs or animations, according to the type of data to analyse. In presenting some of the functionalities offered by our integrated system, we will refer to the two scenarios illustrated in Section 3.

7.1 Analysing Boats Sailing on the Adriatic Sea

In this section we discuss the TDW model and the visual OLAP functionalities of a prototype developed for the analysis of a trajectory dataset referring to boats sailing on the Adriatic sea (see Section 3.1). In addition to the trajectory dataset (time-stamped positions of the boats), for each boat some further information is available, like the type of boat, the daily amount of fish caught, the fish species and the fishing tool used. As a first step we need to adequately instantiate our framework, i.e., choose the right dimensions and hierarchies, and select suitable measures. Figure 7 presents the resulting conceptual model for the TDW.

The hierarchy for the *spatial* dimension consists of a collection of regular grids of increasing size. For the *temporal* dimension, the hierarchy has as base granularity a one day interval. With respect to the general TDW model in Fig. 1, the object group dimension has been instantiated with different dimensions (*boats*, *tools*, *activities*) each with a distinct hierarchy. The *boat* dimension is associated with a hierarchy identifying at the base level the single boats, that can then be aggregated according to the navy they belong to. The *activity* and *tools* dimensions have a hierarchy consisting only of the base granularity. The *tools* dimension provides information on the fishing gear used by a boat, while the *activity* dimension indicates the activity performed by a boat (e.g., fishing, moving to reach a place, going back to the harbour). The

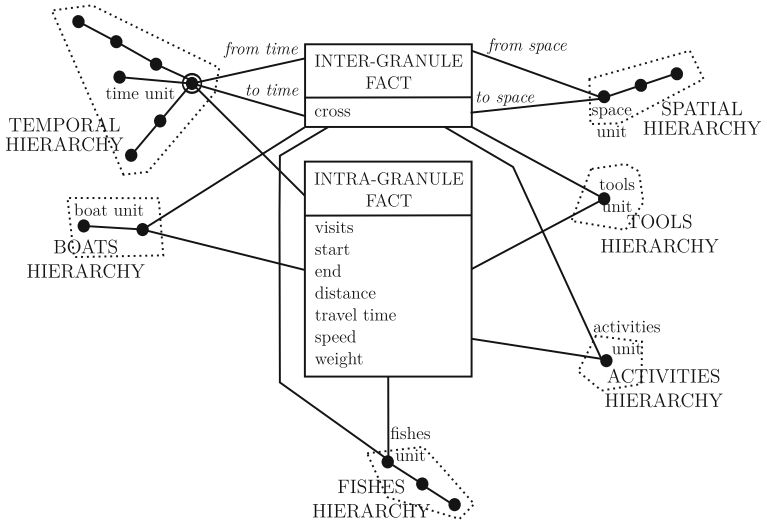


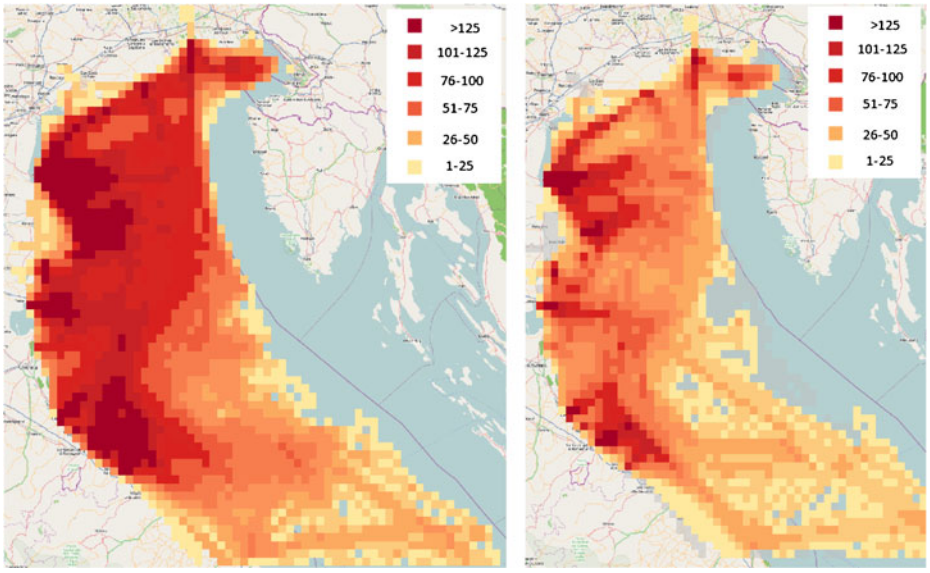
Fig. 7 Actual TDW schema for ships scenario

fishes dimension gives information about the caught fishes. The lowest level of the hierarchy identifies the specific kind of fish. Fishes can be grouped according to their species and in turn according to their family.

Concerning the measures, note that the intra-granule fact table, compared to the general model, contains a new measure *weight*, which indicates the amount in kilograms of fish caught in a granule.

We next present some examples of analyses that can be accomplished by using the TDW prototype. As already mentioned, the environmental scientists were mainly interested in analysing two aspects of the fishing activity: the fishing effort index and the distribution of the species on the sea. In connection to this, we discuss how the TDW can be used for understanding the mobility of the vessels (how the vessels move in the sea), thus assessing its environmental impact.

The maps in Fig. 8 are related to the so called fishing effort index, a value which indicates how much a given area has been exploited by the boats fishing in it. First, the swept area, i.e. the total area of a granule that a boat has used for its fishing activity, is computed as the product between the distance covered by the boat in the granule and the size of the fishing tool (one can think of the tool as a net that occupies a certain area). Then the fishing effort index is calculated as the ratio between the total swept area of boats in the granule and the area of the granule itself. Here, some questions of interest are "What is the fishing effort for a given area during 2007?", "How is this effort distributed during the year?", "What effort is due to boats of a given navy?", "How is the effort affected by the kind of tool used by the boats?". All these questions can be answered by our TDW system, by issuing specific OLAP queries and exploiting the visual analytics tool. For specifying these queries, a combination of interactive visualization and dialogs is used. Thus, areas of interest can be marked on map or selected interactively through simple queries or sophisticated sequences of queries (e.g. selecting areas that have high peaks of activities at morning hours of working days). OLAP-style queries are specified using

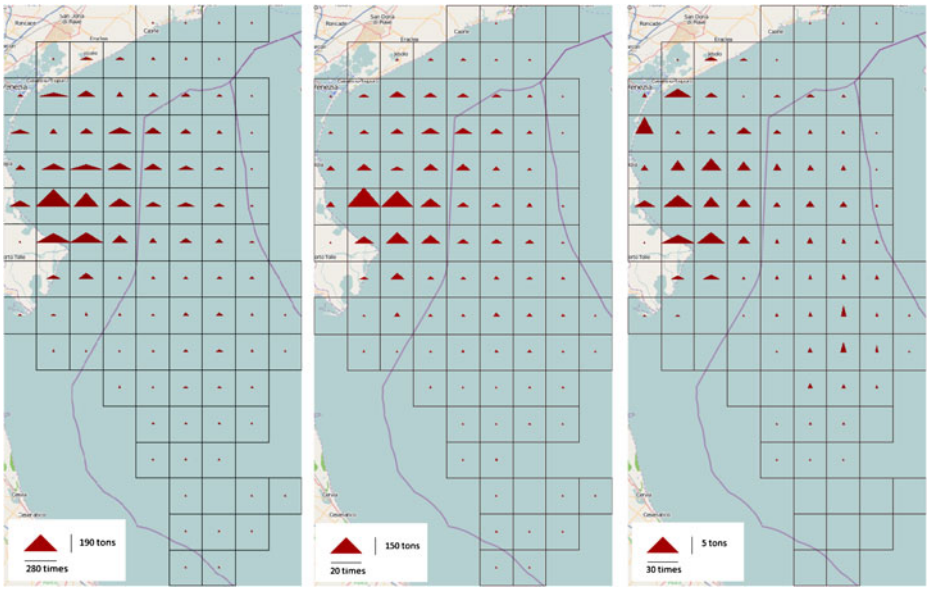


(a) January-September 2007 (b) July-September 2007

Fig. 8 Fishing effort distribution

traditional dialog elements specifying query target and setting restrictions through standard dialog elements (check-boxes for multiple choice options, radio buttons and drop down lists for single choice options, depending on the number of alternatives). A detailed description of the user interface for querying is out of scope of this paper. As an example, Fig. 8a shows the fishing effort in the Northern Adriatic Sea during 2007, at the base spatial granularity. Granules in darker colours are the most exploited. Their index value is greater than 1.5 which means that the corresponding areas have been completely explored more than 1.5 times during the considered period. Note that using a drill-down operation on the temporal dimension, we can inspect the situation at a higher level of detail. For instance, Fig. 8b shows the fishing effort in the trimester July–September of 2007. The fact that it is sensibly reduced with respect to effort in the whole period is somehow expected due to a law which prevents most fishing activities during August.

By combining catches and effort data it is possible to obtain information related to the CPUE (Catch Per Unit Effort), an index commonly used for studying the efficiency of a given fishing technique in a certain environment. In particular, since we can obtain more detailed (and unknown) information than the one given by the aggregated CPUE index, currently used by the research community, the environmental experts appreciated very well the flexibility of our tool. For example, the maps of Fig. 9 use triangular diagrams for indicating the relation between the amount of catches in a given cell (height of the triangle) and the fishing effort for that cell (base of the triangle). In particular, Fig. 9a illustrates the correlation between the two values for the totality of the catches (independently from the fish species), while Figs. 9b and 9c focus, respectively, on anchovies and cuttlefishes. The maps refer to the whole period of interest and spatial cells have been aggregated to form 6×6

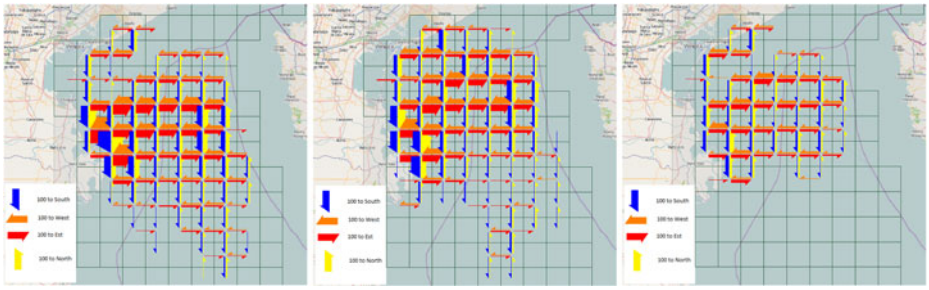


(a) Any fish (b) *Anchovies* (c) *Cuttlefishes*

Fig. 9 Correlation between catches (*height*) and fishing effort (*base*) in the period January–September

nautical miles squares. Cells where the fishing process is not efficient (large effort and few catches) are marked by triangles having a large base and a small height. Vice versa, triangles with small base and large height correspond to the opposite situation, where lots of catches can be made with little effort. This is the case, e.g., for some cells, mostly far from the coast, in Fig. 9c related to cuttlefishes.

The mobility pattern can be understood by aggregating counts of boats moving between adjacent cells at different time periods (Fig. 10). The aggregates are depicted by directional arrows of varying width, proportional to the counts for selected time intervals. We can observe changes in transitions between the cells from one trimester to another.



(a) First trimester (b) Second trimester (c) Third trimester

Fig. 10 Crosses aggregated by trimesters

7.2 Road traffic scenario

In this section we discuss how our TDW model can be used to analyse a road traffic scenario. We focus on a large dataset, collected by GPS-equipped cars moving in the urban area of Milan (Italy) (see Section 3.2). The general TDW model of Fig. 1 when instantiated to this scenario has a pretty complex spatial hierarchy. This is illustrated in Fig. 3b, where the base granule is the segment of the road network. Note that the non spatio-temporal dimension (*object group*) is not present, since the raw data do not include any specific information on the various cars and their owners.

It is important to remark that at the creation phase, the designer can choose the most suitable spatio-temporal hierarchy for his/her purpose, and multiple hierarchies can coexist in our TDW. For example, from a segment, we can go up to a grid cell (a square polygon), to a road (piecewise line), or to a city district (an arbitrary polygon). Of course, the road *segments*, which are the base granules, must be chosen so to be completely included in the ancestors of the four hierarchies illustrated in Fig. 3b. Moreover, relying on the analytical features of V-analytics, our system can also support the designer in the definition of the spatial hierarchy. The designer, for example, can use V-analytics to generate a Voronoi tessellation of the space based on the distribution of the original data [2], as shown in Fig. 12b. This partition has been proved enlightening in order to discover the main directions of the movement of objects.

The flexibility in the definition of the spatio-temporal hierarchy offered by the new TDW model enhances our previous proposal [44], and allows the user to adopt a suitable model of the reality, thus obtaining a much more meaningful visual representation of the information contained in the TDW. Figures 11 and 12 highlight the improvement, from a conceptual and a visual perspective, determined by the use of arbitrarily shaped geometry for the spatial dimension. The images in Fig. 11 visually represent the number of visits to spatial granules during the time interval corresponding to a particular temporal granule. Each image corresponds to a different spatial granularity: in Fig. 11a granules are cells of a regular grid, whereas in Fig. 11b and in Fig. 11c granules are respectively street segments and city districts. The results obtained with a regular grid may be suited for getting an initial overview of the data. However, a more detailed exploration is complicated since the cells do not bear any semantics and do not correspond to the real geographic and topographic properties of the data.

If we are interested in more complex measures, such as the number of crosses of a border between two touching spatial granules, the use of a regular grid may produce a hardly understandable or even misleading result since it greatly distorts the major directions of the movement. Figure 12a shows a set of trajectories from the road traffic dataset, and Fig. 12b illustrates the number of trajectories traversing the border between two adjacent Voronoi regions using arrows. The thicker an arrow is, the higher is the number of trajectories that crossed the border in the corresponding direction. We can easily understand the overall flow of moving objects. On the other end, this is much less evident using a grid as in Fig. 12c. Even if we double the resolution as in Fig. 12d, obtaining grid cells that are significantly smaller than the previous Voronoi regions, it is difficult to grasp major movement flows due to relevant distortions of the prevailing movement directions, which persist despite increasing the resolution.

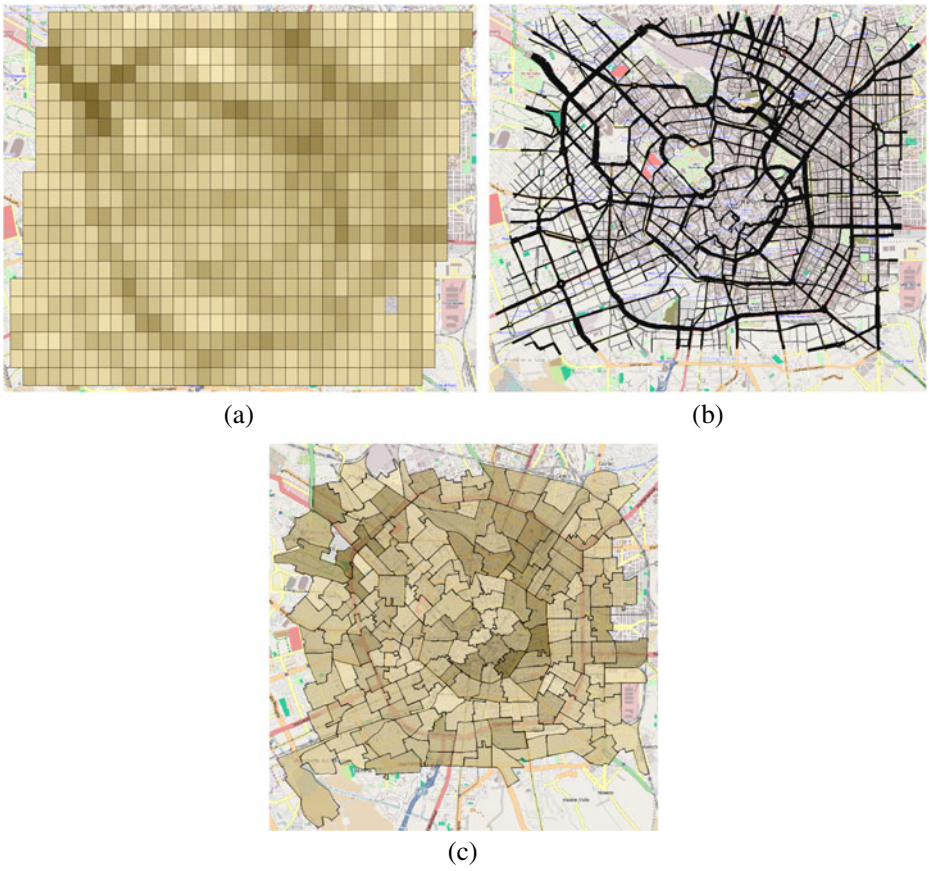


Fig. 11 **a** Grid based spatial dimension, and **b** street segment based spatial dimension with **c** dimensional attribute having polygon spatial type

The proposed procedure of data-driven Voronoi tessellation has been used in a variety of applications related to constrained movement, ranging from vessel traffic described in [7] to indoor movement. In all cases results correctly reflected the network structure, thus enabling further analysis. Respectively, extending the TDW model by a hierarchy of polygon-based aggregations is beneficial for many applications.

Starting from this visualization of the space, one can then decide to explore some measures, which can be visualized according to several methods. The *unclassified choropleth map* technique, for example, fills granules with colour shades so that the degree of darkness is proportional to the value of a selected measure. The *line thickness* visualization technique, draws linear symbols whose thickness is proportional to the value of a given TDW measure. These visualization methods can be used in animated displays, where each frame represents the selected measure in one time interval from the period of interest.

The aggregates obtained by OLAP operations should not be considered as a final product that only needs to be nicely represented for reporting, but rather as

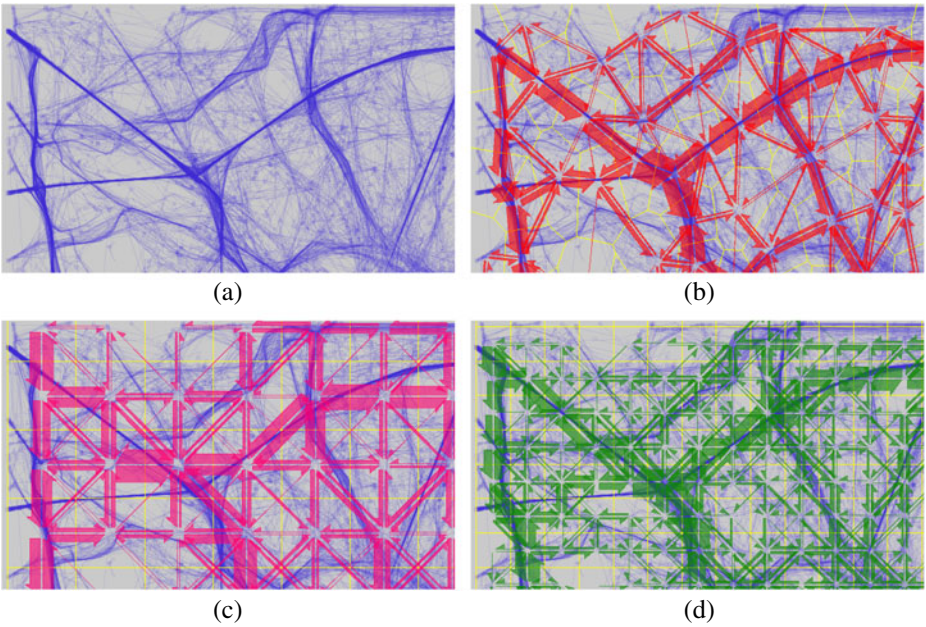
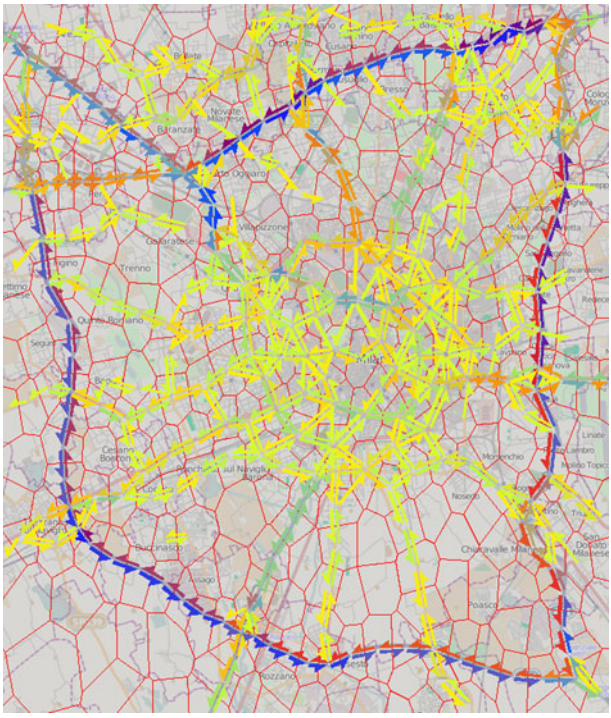


Fig. 12 **a** A set of trajectories, and **b** the corresponding number of trajectories crossing borders between touching areas for Voronoi tessellation and **c**–**d** for regular grids

Fig. 13 Clustering of borders based on cross measure time series



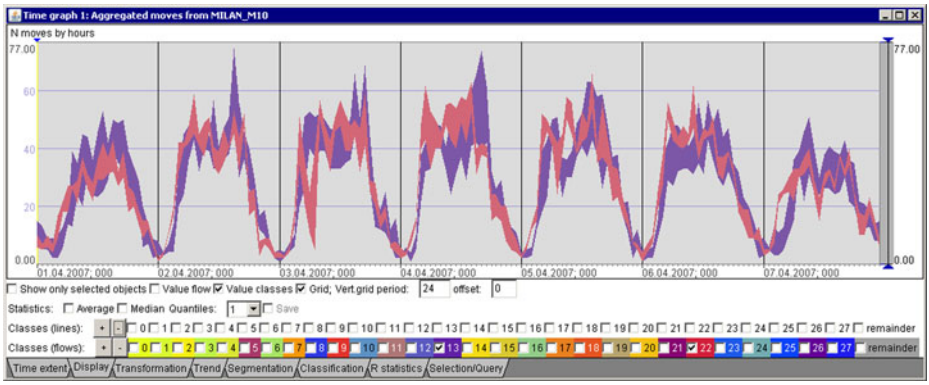


Fig. 14 Variation intervals for the time series belonging to two selected clusters

aggregated data that need to be further explored. The analysis could require the application of additional computational techniques such as clustering and statistical analysis of time series. Using V-analytics, for example, we can analyze measures with the help of clustering on the basis of the time series of the measure values. In Fig. 13 we focused on crosses between adjacent Voronoi regions, considering the flows in opposite directions as two distinct flows. Arrows between areas are coloured according to the similarity of the time series of the cross measure. Distinct colours of the opposite arrows between two areas mean that their temporal variation patterns are rather different. Same or similar colours of successive links going along a road mean that the temporal variation patterns are similar on rather long parts of the road. For example, we can observe in Fig. 13 two clusters (marked by dark blue and dark red colours) that correspond to driving through the belt road in clockwise and anti-clockwise directions, respectively. Figure 14 shows the temporal variation of cross counts between polygons for these two clusters: each colored band (blue and red) indicates the evolution of the maximum (upper bound of the band) and minimum (lower bound) values of the cross measure for all the transitions between polygons belonging to the corresponding cluster. This aggregated representation suggests that there exists a dependency of the count measure from the hour of day and day of week. This dependency, as well as intra-measures dependencies (e.g. dependency of speed on the count of cars) can be further explored and formally modelled using the approach proposed in [6], resulting in a prediction model for road traffic.

8 Trajectory data warehouse implementation

This section provides a brief description of how the proposed conceptual model has been translated into a logical model to be used in the prototypes realized for the case studies.

The prototypes have been implemented using Oracle™ 11 DBMS suite, with the Oracle Spatial extension, needed for the spatial queries used to realize the visual interface described in Section 7.

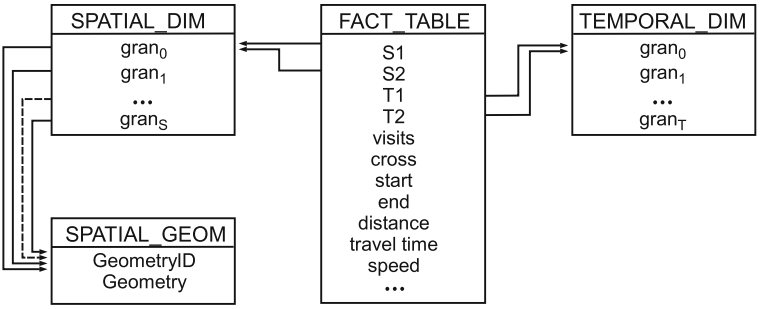


Fig. 15 TDW Logical Model

The main difference between the conceptual model and the logical model concretely used in the implementation is that the latter contains a single fact table, resulting as the union of the intra-granule and inter-granule fact tables. This introduces some level of redundancy, but it eases the management and querying of the data. Figure 15 depicts a simplified version of such a logical model, where only spatial and temporal dimensions are included. Other dimensions could be added by defining new tables with similar schemata.

The *SPATIAL_GEOM* table contains the Oracle *SDO_GEOM*[2] objects representing the spatial granules of the spatial hierarchy. Note that we store spatial granules of any granularity, not only base ones in order to avoid expensive spatial join queries to determine the geometries of coarser spatial granules. This is possible since each hierarchy must be defined in advance when instantiating the TDW conceptual model for a certain scenario. It is worth remarking that only granules containing information are stored in the table during the ETL phase.

The two dimensional tables *SPATIAL_DIM* and *TEMPORAL_DIM* define the hierarchies of the spatial and temporal dimensions respectively. The columns of each table represent the granularities of the corresponding hierarchy, starting from the base granularity, which provides the primary key. Hence each record corresponds to a base granule g identified by $gran_0$, and the columns $gran_i$, with $i > 0$, indicate the (identifiers of the) coarser granules which g belongs to.

The fact table contains the values of the various measures for the base granules. Note that each tuple in the fact table refers to two granules $(T1, S1)$, $(T2, S2)$ (see Fig. 15), and there are two possibilities:

1. when the granules $(T1, S1)$, $(T2, S2)$ coincide, the tuple contains INTRA-GRANULE measures for the granule $(T1, S1)$ and the only INTER-GRANULE measure *Cross* is set to 0;
2. when the granules $(T1, S1)$, $(T2, S2)$ are different, the tuple stores the INTER-GRANULE measure *Cross*, and the INTRA-GRANULE measures are set to 0.

The computation of aggregate values in a roll up operation is accomplished by a standard SQL query. The query is the following, where $gran_s$ and $gran_t$ represent the

[2] http://docs.oracle.com/cd/E11882_01/appdev.112/e11830/sdo_objgeom.htm

granularity of the desired aggregation over space and time, respectively (recall that $gran_0$ is instead the base granularity).

```
SELECT s1.gran_s AS S1,
    s1.gran_s AS S2,
    t1.gran_t AS T1,
    t1.gran_t AS T2,
    SUM(f.Visits) - SUM(f.Cross) AS Visits,
    0 AS Cross,
    SUM(f.Distance) AS Distance
FROM SPATIAL_DIM s1, SPATIAL_DIM s2,
    TEMPORAL_DIM t1, TEMPORAL_DIM t2,
    FACT_TABLE f
WHERE f.S1 = s1.gran_0
    AND f.S2 = s2.gran_0
    AND f.T1 = t1.gran_0
    AND f.T2 = t2.gran_0
    AND s1.gran_s = s2.gran_s
    AND t1.gran_t = t2.gran_t
GROUP BY t1.gran_t,
    s1.gran_s
UNION ALL
SELECT s1.gran_s AS S1,
    s2.gran_s AS S2,
    t1.gran_t AS T1,
    t2.gran_t AS T2,
    0 AS Visits,
    SUM(f.Cross) AS Cross,
    0 AS Distance
FROM SPATIAL_DIM s1, SPATIAL_DIM s2,
    TEMPORAL_DIM t1, TEMPORAL_DIM t2,
    FACT_TABLE f
WHERE f.S1 = s1.gran_0
    AND f.S2 = s2.gran_0
    AND f.T1 = t1.gran_0
    AND f.T2 = t2.gran_0
    AND (s1.gran_s <> s2.gran_s
    OR t1.gran_t <> t2.gran_t)
GROUP BY t1.gran_t,
    t2.gran_t,
    s1.gran_s,
    s2.gran_s
```

The query consists of two parts, combined by a union operator, which computes the aggregate values for intra-granules and inter-granule measures, respectively (for the sake of simplicity, we consider only two intra-granule measures *Visits* and *Distance*). In the two queries, we do not need to distinguish between tuples for intra-granule or inter-granules measures, and we sum up all the available data. In fact,

by construction, in the loading phase the inter-granule measures are set to 0 if the two granules coincide; similarly, the intra-granules measures are set to 0 when the two granules are distinct. Note that since, for each spatial base granule the table *SPATIAL_DIM* stores the coarser spatial granules it belongs to, we do not need to perform spatial queries in order to compute aggregations, but we simply make a `GROUP BY` on the desired spatio-temporal granularity.

9 Conclusion

We have discussed a theoretical and general framework for the design of a TDW, which can be exploited in different application scenarios. In particular, we have introduced an original conceptual model of the TDW, aimed at storing aggregate measures about trajectories. We have formally defined spatial and temporal dimensions, some significant measures and the relative aggregation functions.

With respect to our previous works, the TDW model proposed in this paper supports a flexible discretization of the spatial and temporal domains, along with the associated dimension hierarchies. This allows users to adopt a more suitable model of the reality, by also obtaining a more meaningful visual representation of the information contained in the TDW.

Among the other measures, we have introduced and analysed in depth the aggregation function of measure $\mathscr{V}$, which represents the total number of *visits* in a given spatio-temporal area of several trajectories. We have formally proved that $\mathscr{V}$ can be computed in an exact way by our TDW, and it provides a very good approximation of the presence $\mathscr{P}$, a measure which counts the number of *distinct* trajectories occurring in the same spatio-temporal area. This result is independent of the discretization of the spatio-temporal domains and the specific hierarchies adopted in the TDW according to the application scenario. It is definitively of interest, since the aggregate function for $\mathscr{P}$ is *holistic* [23] and this kind of functions represents a big issue for data warehouse technology.

Finally, by using demonstrators of our TDW framework, devised for two different application scenarios, we have also presented the main features of the Visual OLAP component of the framework.

For the future, we plan to add new complex measures to our TDW. For example, we are currently working on adding *frequent patterns* and *representative trajectories* as measures. Some preliminary results of this effort can be found in [32]. Another interesting future evolution concerns the handling of semantically annotated trajectories, where specific episodes contained in the trajectories are associated with a semantics. Even if this is already possible with the current model, as we did in the vessels scenario to build the activity dimension (fishing, moving, entering/exiting the harbour), an in depth analysis of trajectories behaviour calls for a more complex model.

Acknowledgements This work has been partially supported by the national research project PON "TETRis" (no. PON01_00451), the Marie Curie Project SEEK (no. 295179) and the Cost Action MOVE (no. IC0903). We are grateful to our colleagues of the Department of Environmental Sciences for their support in the analysis of the vessels scenario. We thank the anonymous referees for their useful suggestions and Paolo Baldan for his careful reading of the paper.

A Appendix

This appendix includes the proofs of the propositions presented in Section 5, as well as some formal definitions and lemmas that are only referred inside proofs.

The spatio-temporal operators provided by the MOD allow to load into the TDW the correct values for the measures at the base granularity. Here we prove that they remain correct also when roll-up operations are performed, as stated by Proposition 1. The proof of this proposition is divided into two lemmata, i.e, Lemmas 1 and 4.

We start with a lemma which takes into account the aggregate functions for all the measures except $\mathscr{V}$.

Lemma 1 *Let $\mathscr{H}_{TS}$ be a hierarchy, let T be a set of trajectories and U be an object group. For any $G \in \mathscr{H}_{TS}$ with $G \neq G_\perp$ and $g, g' \in G$ with $g \neq g'$*

$$\mathscr{S}^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} \mathscr{S}^{\mathsf{T}}(g_p, U) \tag{2}$$

$$\mathscr{E}^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} \mathscr{E}^{\mathsf{T}}(g_p, U) \tag{3}$$

$$dist^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} dist^{\mathsf{T}}(g_p, U) \tag{4}$$

$$trav\_t^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} trav\_t^{\mathsf{T}}(g_p, U) \tag{5}$$

$$\mathscr{C}^{\mathsf{T}}(g, g', U) = \Sigma_{g_p \subseteq g, g'_p \subseteq g'} \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) \tag{6}$$

where $g_p, g'_p \in G_p$ with $g_p \neq g'_p$ and $G_p \preceq G$ and $G_p \neq G$.

Proof

$$\begin{aligned}
&\mathscr{S}^{\mathsf{T}}(g, U) = \\
&\qquad \text{by Definition of } \mathscr{S} \\
&= |\{id \mid id \in U \wedge \delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle \wedge s^1_{id} \subseteq g\}| \\
&\qquad \text{by Definition 3, } G_\perp \preceq G_P \text{ and } G_p \preceq G \\
&= |\{id \mid id \in U \wedge \delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle \wedge s^1_{id} \subseteq g_p \\
&\qquad\qquad \textit{for some } g_p \subseteq g\}| \\
&\qquad \text{since } \mathrm{G_p} \text{ is a partition } \mathrm{g_p} \text{ is unique} \\
&= \Sigma_{g_p \subseteq g} |\{id \mid id \in U \wedge \delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle \wedge s^1_{id} \subseteq g_p\}| \\
&\qquad \text{by Definition of } \mathscr{S} \\
&= \Sigma_{g_p \subseteq g'} \mathscr{S}^{\mathsf{T}}(g_p, U)
\end{aligned}$$

The remaining statements are proved in a similar way. □

The case of measure $\mathscr{V}$ is more complex and requires the introduction of an auxiliary measure B, counting the number of times trajectories cross the border of a granule.

Definition 8 Let G be a spatio-temporal granularity and $g \in G$ be a granule, let T be a set of trajectories and let U be an object group. We denote by $B^{\mathsf{T}}(g, U)$ the number of intersections between the trajectories of object group U and the border of the spatio-temporal granule g.

$$B^{\mathsf{T}}(g, U) = \Sigma_{g' \in G, g' \neq g} \left(\mathscr{C}^{\mathsf{T}}(g, g', U) + \mathscr{C}^{\mathsf{T}}(g', g, U) \right)$$

The following lemma states how measure B can be computed by using sub-aggregates of $\mathscr{C}$.

Lemma 2 *Let $\mathscr{H}_{TS}$ be a hierarchy, let T be a set of trajectories and U be an object group. For any $G \in \mathscr{H}_{TS}$ with $G \neq G_\perp$ and $g \in G$*

$$B^{\mathsf{T}}(g, U) = \sum_{g_p \subseteq g,\, g'_p \not\subseteq g} \left(\mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) + \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U) \right)$$

where $g_p, g'_p \in G_p$ with $G_p \preceq G$ and $G_p \neq G$.

Proof

$$\begin{aligned}
&B^{\mathsf{T}}(g, U) = \\
&\qquad \text{by definition of } B \\
&= \textstyle\sum_{g' \in G,\, g' \neq g} (\mathscr{C}^{\mathsf{T}}(g, g', U) + \mathscr{C}^{\mathsf{T}}(g', g, U)) \\
&\qquad \text{by statement (6) of Lemma 1} \\
&= \textstyle\sum_{g' \in G,\, g' \neq g} (\sum_{g_p \subseteq g, g'_p \subseteq g'} \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) + \\
&\qquad\qquad \textstyle\sum_{g_p \subseteq g, g'_p \subseteq g'} \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U)) \\
&= \textstyle\sum_{g' \in G,\, g' \neq g} (\sum_{g_p \subseteq g, g'_p \subseteq g'} \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) + \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U)) \\
&\qquad G \text{ is a partition} \\
&= \textstyle\sum_{g_p \subseteq g, g'_p \not\subseteq g} (\mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) + \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U))
\end{aligned}$$

□

The next lemma provides an inductive characterisation of measure B.

Lemma 3 *Let $\mathscr{H}_{TS}$ be a hierarchy, let T be a set of trajectories and U be an object group. For any $G \in \mathscr{H}_{TS}$ with $G \neq G_\perp$ and $g \in G$*

$$B^{\mathsf{T}}(g, U) = \Sigma_{g_p \subseteq g} B^{\mathsf{T}}(g_p, U) - 2\Sigma_{g_p, g'_p \subseteq g, g_p \neq g'_p} \left(\mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) + \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U) \right)$$

where $g_p, g'_p \in G_p$ with $G_p \preceq G$ and $G_p \neq G$.

Proof This property easily follows by definition of B. More in detail, for the sake of simplicity assume $g = g_p \cup g'_p$. Then the border of the granule g consists of the union of the borders of g_p and g'_p minus the common border between the granules g_p and g'_p. As a consequence trajectories crossing the common border remain inside the granule g and they should not be counted in $B^{\mathsf{T}}(g, U)$. In formulas:

$$\begin{aligned} B^{\mathsf{T}}(g, U) = {} & B^{\mathsf{T}}(g_p, U) - \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) - \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U) \\ & + B^{\mathsf{T}}(g'_p, U) - \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U) - \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) \end{aligned}$$

This is exactly the desired result. □

Now we show how the measure $\mathscr{V}$ can be expressed analytically in terms of B, $\mathscr{S}$ and $\mathscr{E}$.

Proposition 4 *Let G be a spatio-temporal granularity, let T be a set of trajectories and let U be an object group. Then for each $g \in G$ the following statement holds:*

$$\mathscr{V}^{\mathsf{T}}(g, U) = \frac{B^{\mathsf{T}}(g, U) + \mathscr{S}^{\mathsf{T}}(g, U) + \mathscr{E}^{\mathsf{T}}(g, U)}{2}$$

Proof We observe that, as obvious from its definition, $\mathscr{V}$ can be computed by summing up the contributions given to such a measure separately by each trajectory. More precisely, let g be a granule and $\mathsf{T} = \{\mathsf{T}_{id}\}_{id \in \mathscr{I}}$ a set of trajectories. Then $\mathscr{V}^{\mathsf{T}}(g, U) = \sum_{id \in \mathscr{I}} \mathscr{V}^{\{\mathsf{T}_{id}\}}(g, U)$.

Therefore, we can prove the proposition for a single trajectory T_{id} and thesis will trivially extend to a generic set of trajectories.

Let $\delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle$, we prove the thesis by induction on n (the number of sub-trajectories of the trajectory decomposition).

[$n = 1$] In this case $\delta(\mathsf{T}_{id}) = \langle s^1_{id} \rangle$. If $s^1_{id} \subseteq g$, then $\mathscr{V}^{\mathsf{T}}(g, U) = 1$, $B^{\mathsf{T}}(g, U) = 0$, $\mathscr{S}^{\mathsf{T}}(g, U) = 1$, and $\mathscr{E}^{\mathsf{T}}(g, U) = 1$. Thus the thesis holds.
If, instead, $s^1_{id} \not\subseteq g$, then $\mathscr{V}^{\mathsf{T}}(g, U) = 0$, $B^{\mathsf{T}}(g, U) = 0$, $\mathscr{S}^{\mathsf{T}}(g, U) = 0$, and $\mathscr{E}^{\mathsf{T}}(g, U) = 0$ and the thesis holds.
Notice that in both cases the measure B is equal to 0 since at least two sub-trajectories are necessary to have $B > 0$ because each sub-trajectory belongs to exactly one granule.

[$n \Rightarrow n + 1$] In this case $\delta(\mathsf{T}_{id}) = \langle s^1_{id}, \ldots, s^n_{id}, s^{n+1}_{id} \rangle$. Let $\delta(\mathsf{T}'_{id}) = \langle s^1_{id}, \ldots, s^n_{id} \rangle$. We consider four cases:

[case $s^n_{id} \subseteq g$ and $s^{n+1}_{id} \subseteq g$] By definition $\mathscr{V}^{\mathsf{T}}(g, U) = \mathscr{V}^{\mathsf{T}'}(g, U)$, $\mathscr{S}^{\mathsf{T}}(g, U) = \mathscr{S}^{\mathsf{T}'}(g, U)$, $\mathscr{E}^{\mathsf{T}}(g, U) = \mathscr{E}^{\mathsf{T}'}(g, U)$, and $B^{\mathsf{T}}(g, U) = B^{\mathsf{T}'}(g, U)$. Thus, the thesis holds by inductive hypothesis.

[case $s_{id}^{n} \not\subseteq g$ and $s_{id}^{n+1} \subseteq g$] By definition $\mathscr{V}^{\mathsf{T}}(g,U) = \mathscr{V}^{\mathsf{T}'}(g,U)+1$, $\mathscr{S}^{\mathsf{T}}(g,U) = \mathscr{S}^{\mathsf{T}'}(g,U)$, $\mathscr{E}^{\mathsf{T}}(g,U) = \mathscr{E}^{\mathsf{T}'}(g,U)+1$, and $B^{\mathsf{T}}(g,U) = B^{\mathsf{T}'}(g,U)+1$. Thus by using the inductive hypothesis

$$\begin{aligned}\mathscr{V}^{\mathsf{T}}(g,U) &= \mathscr{V}^{\mathsf{T}'}(g,U)+1\\ &= \frac{B^{\mathsf{T}'}(g,U)+\mathscr{S}^{\mathsf{T}'}(g,U)+\mathscr{E}^{\mathsf{T}'}(g,U)}{2}+1\\ &= \frac{B^{\mathsf{T}'}(g,U)+1+\mathscr{S}^{\mathsf{T}'}(g,U)+\mathscr{E}^{\mathsf{T}'}(g,U)+1}{2}\\ &= \frac{B^{\mathsf{T}}(g,U)+\mathscr{S}^{\mathsf{T}}(g,U)+\mathscr{E}^{\mathsf{T}}(g,U)}{2}\end{aligned}$$

[case $s_{id}^{n} \subseteq g$ and $s_{id}^{n+1} \not\subseteq g$] By definition $\mathscr{V}^{\mathsf{T}}(g,U) = \mathscr{V}^{\mathsf{T}'}(g,U)$, $\mathscr{S}^{\mathsf{T}}(g,U) = \mathscr{S}^{\mathsf{T}'}(g,U)$, $\mathscr{E}^{\mathsf{T}}(g,U) = \mathscr{E}^{\mathsf{T}'}(g,U)-1$, and $B^{\mathsf{T}}(g,U) = B^{\mathsf{T}'}(g,U)+1$. As in the previous case we can conclude.

[case $s_{id}^{n} \not\subseteq g$ and $s_{id}^{n+1} \not\subseteq g$] By definition $\mathscr{V}^{\mathsf{T}}(g,U) = \mathscr{V}^{\mathsf{T}'}(g,U)$, $\mathscr{S}^{\mathsf{T}}(g,U) = \mathscr{S}^{\mathsf{T}'}(g,U)$, $\mathscr{E}^{\mathsf{T}}(g,U) = \mathscr{E}^{\mathsf{T}'}(g,U)$, and $B^{\mathsf{T}}(g,U) = B^{\mathsf{T}'}(g,,U)$. Thus the thesis holds by inductive hypothesis. □

The following lemma concludes the proof for measure $\mathscr{V}$.

Lemma 4 *Let $\mathscr{H}_{TS}$ be a hierarchy, let T be a set of trajectories and U be an object group. For any $G \in \mathscr{H}_{TS}$ with $G \neq G_{\perp}$ and $g \in G$*

$$\mathscr{V}^{\mathsf{T}}(g,U) = \Sigma_{g_p \subseteq g}\left(\mathscr{V}^{\mathsf{T}}(g_p,U) - \Sigma_{g'_p \subseteq g}\mathscr{C}^{\mathsf{T}}(g_p,g'_p,U)\right)$$

where $g_p \in G_p$ with $G_p \preceq G$ and $G_p \neq G$.

Proof For the sake of simplicity, let $g = g_p \cup g'_p$ with $g_p, g'_p \in G_p$ and $g_p \neq g'_p$.

$$\mathscr{V}^{\mathsf{T}}(g,U) =$$

by Proposition 4

$$= \frac{B^{\mathsf{T}}(g,U)+\mathscr{S}^{\mathsf{T}}(g,U)+\mathscr{E}^{\mathsf{T}}(g,U)}{2}$$

by Lemma 3 and $g = g_p \cup g'_p$

$$= \frac{B^{\mathsf{T}}(g_p,U)+B^{\mathsf{T}}(g'_p,U)-2\left(\mathscr{C}^{\mathsf{T}}(g_p,g'_p,U)+\mathscr{C}^{\mathsf{T}}(g'_p,g_p,U)\right)+\mathscr{S}^{\mathsf{T}}(g,U)+\mathscr{E}^{\mathsf{T}}(g,U)}{2}$$

By Lemma 1, we have

$$\mathscr{S}^{\mathsf{T}}(g,U) = \mathscr{S}^{\mathsf{T}}(g_p,U)+\mathscr{S}^{\mathsf{T}}(g'_p,U)$$

and

$$\mathscr{E}^{\mathsf{T}}(g, U) = \mathscr{E}^{\mathsf{T}}(g_p, U) + \mathscr{E}^{\mathsf{T}}(g'_p, U)$$

By Proposition 4

$$\mathscr{V}^{\mathsf{T}}(g_p, U) = \frac{B^{\mathsf{T}}(g_p, U) + \mathscr{S}^{\mathsf{T}}(g_p, U) + \mathscr{E}^{\mathsf{T}}(g_p, U)}{2}$$

and

$$\mathscr{V}^{\mathsf{T}}(g'_p, U) = \frac{B^{\mathsf{T}}(g'_p, U) + \mathscr{S}^{\mathsf{T}}(g'_p, U) + \mathscr{E}^{\mathsf{T}}(g'_p, U)}{2}$$

Hence we can conclude

$$\mathscr{V}^{\mathsf{T}}(g, U) = \mathscr{V}^{\mathsf{T}}(g_p, U) + \mathscr{V}^{\mathsf{T}}(g'_p, U) - \mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) - \mathscr{C}^{\mathsf{T}}(g'_p, g_p, U)$$

This is the thesis since $\mathscr{C}^{\mathsf{T}}(g, g, U) = 0$ for any granule g. □

The join of Lemmas 1 and 4 is exactly Proposition 1.

Proposition 1 *Let $\mathscr{H}_{TS}$ be a hierarchy, let T be a set of trajectories and U be an object group. For any $G \in \mathscr{H}_{TS}$ with $G \neq G_\perp$, $g, g' \in G$ with $g \neq g'$*

$$\begin{aligned}
\mathscr{V}^{\mathsf{T}}(g, U) &= \Sigma_{g_p \subseteq g}\left(\mathscr{V}^{\mathsf{T}}(g_p, U) - \Sigma_{g'_p \subseteq g}\mathscr{C}^{\mathsf{T}}(g_p, g'_p, U)\right) \\
\mathscr{S}^{\mathsf{T}}(g, U) &= \Sigma_{g_p \subseteq g}\mathscr{S}^{\mathsf{T}}(g_p, U) \\
\mathscr{E}^{\mathsf{T}}(g, U) &= \Sigma_{g_p \subseteq g}\mathscr{E}^{\mathsf{T}}(g_p, U) \\
\mathscr{C}^{\mathsf{T}}(g, g', U) &= \Sigma_{g_p \subseteq g, g'_p \subseteq g'}\mathscr{C}^{\mathsf{T}}(g_p, g'_p, U) \\
dist^{\mathsf{T}}(g, U) &= \Sigma_{g_p \subseteq g} dist^{\mathsf{T}}(g_p, U) \\
trav_t^{\mathsf{T}}(g, U) &= \Sigma_{g_p \subseteq g} trav_t^{\mathsf{T}}(g_p, U)
\end{aligned}$$

where $g_p, g'_p \in G_p$ with $g_p \neq g'_p$ and G_P is a predecessor of G, i.e., $G_p \preceq G$ and $G_p \neq G$.

We finally prove the assertion regarding the relation between visits and presence.

Proposition 3 *Let G be a spatio-temporal granularity and $g \in G$, let T be a set of trajectories and let U be an object group, then*

1. *if each trajectory visits g at most once then $\mathscr{P}^{\mathsf{T}}(g, U) = \mathscr{V}^{\mathsf{T}}(g, U)$*
2. *if $B^{\mathsf{T}}(g, U) = 0$ then $\mathscr{P}^{\mathsf{T}}(g, U) = \mathscr{V}^{\mathsf{T}}(g, U)$.*

Proof The first statement is a straightforward consequence of definitions $\mathscr{V}$ and $\mathscr{P}$.

In order to prove the second statement observe that the hypothesis $B^{\mathsf{T}}(g, U) = 0$ means that all the trajectories are completely contained in a granule g. Hence $\mathcal{S}^{\mathsf{T}}(g, U) = \mathcal{E}^{\mathsf{T}}(g, U) = \mathscr{P}^{\mathsf{T}}(g, U)$. Then, by Proposition 4, we have that

$$\mathscr{V}^{\mathsf{T}}(g, U) = \frac{\mathscr{S}^{\mathsf{T}}(g, U) + \mathscr{E}^{\mathsf{T}}(g, U)}{2} = \frac{2\mathscr{S}^{\mathsf{T}}(g, U)}{2} = \mathscr{P}^{\mathsf{T}}(g, U)$$

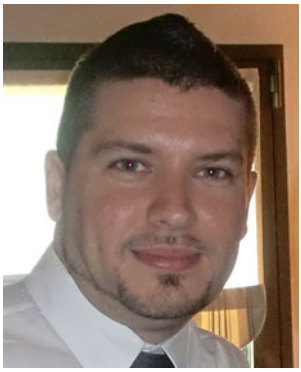
□

References

1. Allen JF (1984) A general model of action and time. Artif Intell 23:123–154
2. Andrienko G, Andrienko N (2008) Spatio-temporal aggregation for visual analysis of movements. In: Proceedings of VAST. IEEE. pp 51–58
3. Andrienko G, Andrienko N (2010) A general framework for using aggregation in visual exploration of movement data. Cartogr J 47(1):22–40
4. Andrienko G, Andrienko N, Wrobel S (2007) Visual analytics tools for analysis of movement data. ACM SIGKDD Explor 9(2):28–46
5. Andrienko N, Andrienko G (2011) Spatial generalization and aggregation of massive movement data. IEEE Trans Vis Comput Graph 17(2):205–219
6. Andrienko N, Andrienko G (2013) A visual analytics framework for spatio-temporal analysis and modelling. Data Min Knowl Disc 27(1):55–83
7. Andrienko N, Andrienko G (2013) Visual analytics of movement: an overview of methods, tools and procedures. Inf Vis 12(1):3–24
8. Bimonte S, Miquel M (2010) When spatial analysis meets OLAP: multidimensional model and operators. IJDWM 6(4):33–60
9. Brakatsoulas S, Pfoser D, Salas R, Wenk C (2005) On map-matching vehicle tracking data. In: Proceedings of VLDB, pp 853–864
10. Brillinger D, Preisler H, Ager A, Kie K (2004) An exploratory data analysis (EDA) of the paths of moving animals. J Stat Plan Infer 122(2):43–63
11. Cao H, Wolfson O, Trajcevski G (2006) Spatio-temporal data reduction with deterministic error bounds. VLDB J 15(3):211–228
12. Cudré-Mauroux P, Wu E, Madden S (2010) TrajStore: an adaptive storage system for very large trajectory data sets. In: Proceedings of ICDE, pp 109–120
13. Dykes JA, Mountain DM (2003) Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. Comput Stat Data Anal 43(4):581–603
14. Egenhofer MJ (1994) Topological relations between regions with holes. Int J GIS 8:129–142
15. Eick SG (2000) Visualizing multi-dimensional data. SIGGRAPH Comput Graph 34:61–67
16. Erwig M, Güting RH, Schneider M, Vazirgiannis M (1999) Spatio-temporal data types: an approach to modeling and querying moving objects in databases. GeoInformatica 3(3):269–296
17. Forer P, Huisman O (2000) Information, place and cyberspace: issues in accessibility, chap. time and sequencing: substitution at the physical/virtual interface. Springer Verlag, Heidelberg, pp 73–90
18. Fredrikson A, North C, Plaisant C, Shneiderman B (1999) Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data. In: Proceedings of workshop on new paradigms in information visualization and manipulation, pp 26–34
19. European project IST-6FP-014915 GeoPKDD Geographic privacy-aware knowledge discovery and delivery (GeoPKDD) (web site: http://www.geopkdd.eu)
20. Golfarelli M, Maio D, Rizzi S (1998) The dimensional fact model: a conceptual model for data warehouses. Int J Coop Inf Syst 7(2–3):215–247
21. Gómez LI, Haesevoets S, Kuijpers B, Vaisman AA (2009) Spatial aggregation: data model and implementation. Inf Syst 34(6):551–576
22. Gómez LI, Kuijpers B, Vaisman AA (2011) A data model and query language for spatio-temporal decision support. GeoInformatica 15(3):455–496
23. Gray J, Chaudhuri S, Bosworth A, Layman A, Reichart D, Venkatrao M, Pellow F, Pirahesh H (1997) Data cube: a relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Min Knowl Disc 1(1):29–53

24. Guo D (2007) Visual analytics of spatial interaction patterns for pandemic decision support. Int J Geograph Inf Sci 21:859–877
25. Güting RH, de Almeida VT, Ding Z (2006) Modeling and querying moving objects in networks. VLDB J 15(2):165–190
26. Güting RH, Behr T, Düntgen C (2010) SECONDO: a platform for moving objects database research and for publishing and integrating research implementations. IEEE Data Eng Bull 33(2):56–63
27. Güting RH, Schneider M (2005) Moving Objects Databases. Morgan Kaufman, San Mateo
28. Han J, Stefanovic N, Kopersky K (1998) Selective materialization: an efficient method for spatial data cube construction. In: Proceedings of PAKDD, pp 144–158
29. Jensen CS, Kligys A, Pedersen TB, Timko I (2004) Multidimensional data modeling for location-based services. VLDB J 13(1):1–21
30. Jensen CS, Lu H, Yang B (2009) Indexing the trajectories of moving objects in symbolic indoor space. In: Proceedings of SSTD, pp 208–227
31. Keim D, Kohlhammer J, Ellis G, Mansmann F (eds) (2010) Mastering the information age solving problems with visual analytics, chap. data mining, pp 39–56. Eurographics Association
32. Leonardi L, Orlando S, Raffaetà A, Roncato A, Silvestri C (2009) Frequent spatio-temporal patterns in trajectory data warehouses. In: Proceedings of SAC. ACM, pp 1433–1440
33. Liu K, Deng K, Ding Z, Li M, Zhou X (2009) MOIR/MT: monitoring large-scale road network traffic in real-time. PVLDB 2(2):1538–1541
34. Malinowski E, Zimányi E (2004) Representing spatiality in a conceptual multidimensional model. In: Proceedings of GIS, pp 12–22
35. Malinowski E, Zimányi E (2007) Logical representation of a conceptual model for spatial data warehouses. GeoInformatica 11(4):431–457
36. Marketos G, Frentzos E, Ntoutsi I, Pelekis N, Raffaetà A, Theodoridis Y (2008) Building Real World Trajectory Warehouses. In: Proceedings of MobiDE, pp 8–15
37. Orlando S, Orsini R, Raffaetà A, Roncato A, Silvestri C (2007) Trajectory data warehouses: design and implementation issues. J Comput Sci Eng 1(2):240–261
38. Papadias D, Tao Y, Kalnis P, Zhang J (2002) Indexing spatio-temporal data warehouses. In: Proceedings of ICDE, pp 166–175
39. Papadias D, Zhang J, Mamoulis N, Tao Y (2003) Query processing in spatial network databases. In: Proceedings of VLDB, pp 802–813
40. Pedersen T, Tryfona N (2001) Pre-aggregation in spatial data warehouses. In: Proceedings of SSTD, LNCS, vol 2121, pp 460–480
41. Pelekis N, Theodoridis Y (2006) Boosting location-based services with a moving object database engine. In: Proceedings of MobiDE, pp 3–10
42. Pfoser D, Jensen CS (2005) Trajectory indexing using movement constraints. GeoInformatica 9(2):93–115
43. Popa IS, Zeitouni K, Oria V, Barth D, Vial S (2011) Indexing in-network trajectory flows. VLDB J 20(5):643–669
44. Raffaetà A, Leonardi L, Marketos G, Andrienko G, Andrienko N, Frentzos E, Giatrakos N, Orlando S, Pelekis N, Roncato A, Silvestri C (2011) Visual mobility analysis using T-warehouse. IJDWM 7(1):1–23
45. Sakr M, Andrienko G, Behr T, Andrienko N, Güting RH, Hurter C (2011) Exploring spatiotemporal patterns by integrating visual analytics with a moving objects database system. In: Proceedings of GIS, pp 505–508
46. Sakr MA, Güting RH (2011) Spatiotemporal pattern queries. GeoInformatica 15(3):497–540
47. Siqueira TLL, de Aguiar Ciferri CD, Times VC, Ciferri RR (2012) The SB-index and the HSB-index: efficient indices for spatial data warehouses. GeoInformatica 16(1):165–205
48. Stolte C, Tang D, Hanrahan P (2002) Polaris: A system for query, analysis, and visualization of multidimensional relational databases. IEEE Trans Vis Comput Graph 8:52–65
49. Sweeney L (2002) Achieving k-anonymity privacy protection using generalization and suppression. Int J Uncertain Fuzz Knowl-Based Syst 10:571–588
50. Tao Y, Kollios G, Considine J, Li F, Papadias D (2004) Spatio-temporal aggregation using sketches. In: Proceedings of ICDE, pp 214–225
51. Tao Y, Papadias D (2005) Historical spatio-temporal aggregation. ACM TOIS 23:61–102
52. Timko I, Pedersen TB (2004) Capturing complex multidimensional data in location-based data warehouses. In: Proceedings of GIS, pp 147–156
53. Veilleux JP, Lambert M, Santerre R, Bédard Y (2004) Utilisation du systéme de positionnement par satellites (GPS) et des outils d'exploration et d'analyse SOLAP pour l'évaluation et le suivi de sportifs de haut niveau. In: Colloque Géomatique 2004 – Un choix stratégique!

54. Wan T, Zeitouni K, Meng X (2007) An OLAP system for network-constrained moving objects. In: Proceedings of SAC, pp 13–18
55. Wolfson O, Xu B, Chamberlain S, Jiang L (1998) Moving objects databases: issues and solutions. In: Proceedings of SSDBM, pp 111–122



Luca Leonardi obtained his Ph.D from Ca' Foscari University of Venice (Italy), supervised by prof. Alessandra Raffaetà. In 2006 he got a Bachelor Degree in Computer Science at University Ca' Foscari - Venezia. In 2008 he took his Master Degree in Computer Science at the same university. His research interests include mining spatio-temporal data, trajectory warehousing and visual OLAP analysis.

Salvatore Orlando (Laurea/MSc. (summa cum laude,1985) and PhD (1991) in Computer Science, University of Pisa) has been an associate professor of Computer Science at Ca' Foscari University of Venice since August 2000. His research interests include the design of scalable algorithms for various data mining and knowledge discovery problems, distributed and P2P systems for information retrieval, parallel/distributed systems and programming environments. He published over 100 papers in international journals and conference/workshop proceedings. He co-chaired conferences (EuroPVM/MPI, SASO), conference tracks, and workshops. He served on the program committees of international conferences, among which the premier conferences on data mining run by ACM, IEEE, SIAM (KDD, ICDM, SDM), and many others, such as ECML/PKDD, Europar, INFOSCALE, CCGRid, SASO.

Alessandra Raffaetà is an assistant professor at Ca' Foscari University of Venice. She graduated in Computer Science (summa cum laude) in 1994 and she took a PhD in Computer Science in 2000 from the University of Pisa. Her research interests include Data warehouses, GISs, spatio-temporal reasoning, design and formal semantics of programming languages and constraint logic programming. She participated to several national and international research projects and she has published over 40 papers on international journals and conferences. She was co-chair of the workshop Complex Reasoning on Geographical Data (2001) and she was member of the program committee of several international workshops.

Alessandro Roncato is an assistant professor of Computer Science at Ca' Foscari University of Venice. He graduated (summa cum laude) in Computer Science in 1989, from the University of Udine, and he took a PhD in Computer Science from the University of Pisa in 1995. His main research interests include distributed computing, spatio-temporal Data Warehouses and data mining. He is currently working on the implementation of a data warehouse for trajectories of moving objects and on the design of adequate visual OLAP operations.

Claudio Silvestri is an assistant professor of Computer Science at Ca' Foscari University of Venice, and was formerly a research fellow at the University of Milan (Italy) and at Istituto di Scienza e Tecnologie dell'Informazione - Consiglio Nazionale delle Ricerche in Pisa (Italy). His research focuses on security and privacy in mobile computing, spatio-temporal data warehouses, data mining algorithms, and high performance computing. Claudio Silvestri has a PhD in Computer Science from the Università Ca' Foscari Venezia. Contact him at silvestri@unive.it.

Gennady Andrienko is a lead scientist responsible for the visual analytics research at the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS). He co-authored the monograph 'Exploratory Analysis of Spatial and Temporal Data' (Springer, 2006), 60+ peer-reviewed journal papers, 20+ book chapters and more than 100 conference papers. Since 2007, Gennady Andrienko is chairing the ICA Commission on GeoVisualization. He co-organized scientific events on visual analytics, geovisualization and visual data mining, and co-edited 10 special issues of journals.

Natalia Andrienko has been working at GMD, now Fraunhofer IAIS, since 1997. Since 2007, she is a lead scientist responsible for the visual analytics research. She co-authored the mono-graph 'Exploratory Analysis of Spatial and Temporal Data', over 60 peer-reviewed journal papers, over 20 book chapters and more than 100 conference papers. She received best paper awards at AGILE 2006 and IEEE VAST 2011 and 2012 conferences, best poster awards at AGILE 2007 and ACM GIS 2011, and VAST challenge award 2008.