Reasoning about Space and Time: Moving towards a Theory of Granularities

João Moura-Pires[1], Ricardo Almeida Silva[1], and Maribel Yasmina Santos[2]

[1] CENTRIA, Faculty of Science and Technology,
New University of Lisbon, Lisbon, Portugal
jmp@fct.unl.pt, ricardofcsasilva@gmail.com
[2] ALGORITMI Research Centre, University of Minho, Portugal
maribel@dsi.uminho.pt

Abstract. Nowadays, the massive amount of spatio-temporal data available exceeds the human capability to absorb them (i.e., to achieve insights). A possible approach to address this issue is through less detailed representations of phenomena so that the data complexity can be decreased making easier for the users to achieve meaningful insights. In this paper, we discuss the state of the art of modeling spatio-temporal phenomena at different levels of detail (LoDs). We found that granularities play an important role to hold spatio-temporal data at different LoDs. A novel granularity framework is proposed, allowing the definition of a granularity over any domain (including spatial and temporal granularities) as well as it allows transposing knowledge from the original domains to granularities (i.e., known relationships and its properties on the domain). Finally, a granularities-based model is proposed, based on the proposed granularity framework, for dealing and relate different LoDs of spatio-temporal data.

Keywords: spatial-temporal data, multigranularity.

1 Context and Motivation

Spatio-temporal data are being gathered at high levels of detail (LoDs) either from a spatial or temporal perspective, resulting into massive volumes of data to be processed, which can be further analyzed by users. Several examples can be found like spatio-temporal data collected from network sensors, remote sensing imagery; spatio-temporal data resulting from the usage of mobile devices (e.g., twitter, foursquare); spatio-temporal data from monitoring sensors used in marine navigation (e.g., Automatic Identification System) or sensors embedded in vehicles (e.g., Tom Tom).

During the analytical activities performed on massive amounts of spatio-temporal data several problems emerge either from the computer or the human viewpoint [1]. From the computer perspective, the data processing becomes computationally very demanding, and not well-suited for the interactive exploration of large volumes of data as the Visual Analytics stands for [1], i.e., a smart combination of automatic algorithms and interactive visualization. One can make effective use of distributed

B. Murgante et al. (Eds.): ICCSA 2014, Part I, LNCS 8579, pp. 328–343, 2014.

and parallel approaches. However solving the efficiency problems is unlikely enough to develop suitable environments for the analysis of massive spatio-temporal data, as the human issues remains to be solved.

Our (as humans) capability to analyze and exploit spatio-temporal data was largely outpaced [1]. Spatio-temporal data are frequently referred as complex data as their characterizing attributes change over time or establish several relationships or interactions with the surrounding environment. The massive amount of data which embodies spatio-temporal dynamism with multivariate connections, containing implicit structures, relationships and interactions, makes the data analyzing process a very challenging task for a human. This issue is related with the Law of Incompatibility stated by Lotfi A. Zadeh [2] "*As complexity rises, precise statements lose meaning and meaningful statements lose precision*". This tradeoff can also be observed in the analysis of massive amounts of spatio-temporal data. Analyzing such data at fine LoDs hardly bring meaningful insights, while at coarse LoDs can provide meaningful knowledge as long as the amount and complexity of data are reduced. In short, data insight can be improved if loss of precision is performed.

Suppose that we are analyzing spatio-temporal data about crimes. In this case, we are interested in analyzing the tendency about where and what time of the day the crimes occurred. For this purpose, we wouldn't probably need to have the LoD of data at meters and seconds regarding the spatial and temporal components, respectively. However, having the spatial component at the provinces level and the time component at day's sub-units (morning, afternoon, evening, night), the previous analytical requirements are met without the need to process data at higher LoDs.

Looking at spatio-temporal data, both temporal and spatial characteristics of data can be expressed at different LoDs using granularities that can range from seconds to months or years, or from points to regions. From our perspective, according to the spatio-temporal phenomenon and analytical goals, different LoDs can be proper to get meaningful insights. In order to provide an instrument capable to reason and relate different LoDs, the contributions of this work may be summarized as follows.

Primarily, we provide a comprehensive study about the ability to model spatio-temporal phenomena at different LoDs, presented in Section 2. Then, a new granularity framework is proposed giving the possibility to define a granularity over any domain in a set of granules disjoint from each other. This framework is more general than existing proposals, in fact, many proposals for spatial and temporal granularities are particular cases of our framework. Moreover, it allows transposing knowledge from the original domains to granularities (i.e., known relationships and its properties on the domain). This way, over a domain, we can specify what kind of properties we intend for certain granularities. Based on the previous framework, presented in Section 3.1, a granularities-based model is proposed that deals with phenomena at different LoDs. To the best of our knowledge, the model distinguishes itself from others works, once every feature of a phenomenon is described by a granule. The model is presented in Section 3.2.

Through the proposed model, we aim to achieve less detailed representations of massive spatio-temporal phenomena in order to provide meaningful insights to the users. Finally, the remarks about the presented work and directions for future work are given in Section 4.

2 Related Work

To model spatio-temporal phenomena at multiple LoDs, multiscale spatio-temporal models have been investigated, proposed by different researchers, with different backgrounds and purposes, using different terminologies like multirepresentation, multiresolution or multigranularity.

2.1 Multirepresentation Approaches

Multirepresentation provides different point of views from a spatio-temporal phenomenon allowing the observation of the same geographical space and/or interval of time, from different perspectives. For example, we can have a representation of a country in terms of unemployment and another representation of the same country in terms of its average temperatures, for a certain time period. In general, the approaches denoted by multirepresentation are based on extensions of the ER (Entity-Relationship) and UML (Unified Modeling Language) models in order to incorporate spatial and temporal features in the database modeling with different LoDs [3]. Several data models, each one with specific concepts, have been proposed in the literature. In [3] a survey about multirepresentation modeling is given in which three requirements are presented that should be verified in a multirepresentation approach. Firstly, a model should allow one to characterize the same object using different sets of attributes, or/and with different domain values. Secondly, a model should allow mapping one object to several objects or two different sets of objects. This is particularly useful when we change the spatial level of detail, where objects may disappear and others may be grouped. Thirdly, a model should enable multiple representations of relationships. For instance, two regions might be modeled as spatially adjacent at lower scale but at more precise scale the regions are just near to each other. According to [3], MADS (Modeling Application Data with Spatio-temporal features) [4] is the only model which verifies the three requirements. It supports multiple spatio-temporal representations of a phenomenon mainly through *perceptions*. More particularly, we can assign *perception stamps* to any element of the schema including objects, object' attributes, and relationships. According to the *perception stamp*, we will have access to different spatial representations of objects or relationships, to different domain values of attributes, or even to different attributes.

Among the main drawbacks of multirepresentation is the fact that different LoDs, required by different applications, or the same application at different stages, can vary [5]. Bearing this in mind, easily the task of modeling a real-world phenomenon for which several spatial and/or temporal LoDs are needed can be very challenging. There are no pre-defined operations that take data from one spatial and/or temporal level to another. Everything is defined by the user at the instances level. Despite these drawbacks, it can be advantageous having several pre-computed representations when dealing with massive spatio-temporal datasets.

2.2 Multiresolution Approaches

Unlike the multirepresentation approaches, the multiresolution is essentially focused on the spatial component of the data. Plus, it derives the proper level of detail on demand [5]. Data are stored at the highest level of resolution (or detail) and are dynamically generalized to lower LoDs, using known and pre-defined generalization operations. The generalization of spatial data is a non-trivial task and involves object simplification, which may lead to a change in the object geometry (e.g., a building can be represented by a polygon at a precise resolution, and by a point at a less precise resolution), dimensionality (e.g., at less precise resolutions, a building may be defined using less vertices than it was originally) and existence (e.g., eventually it is not relevant anymore represent that building). More details about generalization operators can be found in [6].

In an early work, Stell and Worboys [7] define resolution or granularity (for these authors are synonyms) as the level of discernibility between elements of a phenomenon that is being represented by the dataset. We can acquire multiple resolutions of a phenomenon through multiple representations, if we consider that the several representations concern the same geographical space and/or interval of time from the same perspective at different resolutions.

Based on the resolution (or granularity) definition, Stell and Worboys [7] define a *stratified map space* which consists in a set of maps representing the same spatial extent at different resolutions (or granularities) related to form a granularity lattice through general conversion operators (generalize and lift operators). Each map holds the same semantic and spatial granularity which corresponds to a database state. Maps are grouped by map spaces, i.e., sets of maps at the same granularity, describing the set of all possible databases states that are instances of some fixed schema. Through this work, the authors do not intend a formalization of the complex process of cartographic generalization, but a framework as basis reasoning on generalized maps.

In [5] it is proposed a multiresolution approach to generalize polygonal data. The spatial generalization happens in a post-query process based on a *scaless data structure*. Regarding the time required to perform such operation it is not clear. The authors make the following statement: *"We found that the overhead of simplify-while-retrieve approach based on the scaleless data structure is significant but not very large"*. The generalization at run-time is important when such process depends on the data achieved on that moment which in turn may vary according to the user interaction like filtering over semantic attributes, spatial filters, and so on. The time required to perform the generalization process can be an issue. In an interactive application, the fast response time in performing such process is crucial for a user, which is an open issue when dealing with massive datasets [8].

Moreover, to the best of our knowledge, the multiresolution approaches do not provide generalization operators that take into account the temporal component. Finally, they are more focused on the map visualization (and the corresponding spatial generalization operators) and less with the computation of data at different LoDs.

2.3 Multigranularity Approach

Granularity involves semantic aspects of data, both in representation in performing granularity transformations, and differs from resolution, which refers to the amount of detail in a representation. In the literature, temporal and spatial granularities were proposed. A temporal granularity, proposed by Bettini et al. [9], is a sequence of temporal granules, each one composed by a set of time instants. For example, December 2014 can be a temporal granule. Consider a time domain T as a set of totally ordered time instants. A temporal granularity G_t is a mapping from an index set (e.g., the natural numbers) to subsets of the time domain. Suppose that, i, k, and j are elements of an index set. A temporal granularity needs to satisfy the following conditions: (*i*) if $i < j$ and $g_t(i)$ and $g_t(j)$ are non-empty, then each element in $g_t(i)$ is less than all the elements in $g_t(j)$; (*ii*) if $i < k < j$ and $g_t(i)$ and $g_t(j)$ are non-empty, then $g_t(k)$ is non-empty. Each non-empty $g_t(i)$ in the above definition is called a temporal granule. These conditions impose the following: temporal granules of the same temporal granularity cannot overlap as well as non-empty temporal granules must preserve the order given by the index set. Moreover, we cannot have an element (from the index set) mapped to the empty set between any two elements mapped to non-empty subsets. Accordingly, *weeks*, *years* are examples of temporal granularities. On the other hand, a spatial granularity G_s is a set of spatial granules, each one being a portion of a spatial domain. Camossi et al. [10] define spatial granularity as a mapping from an index set to subsets of the spatial domain (assumed as 2-dimensional) such that: if $i \neq j$, and $g_s(i)$ and $g_s(j)$ are non-empty then $g_s(i)$ and $g_s(j)$ are disjoint. No order is required among the spatial granules, but two spatial granules of the same granularity cannot overlap. Examples of spatial granularities are: *countries*, *meters*, among others. The spatial granularity definition is further extended [11] in order to represent also the relations between spatial granules (e.g., direction-based relations, distance-based relations).

Granularities can be related through relationships allowing one to compare and relate granules belonging to different granularities, useful to hold spatio-temporal data at different LoDs [12]. Two commonly used relationships between granularities (spatial or temporal) are given. A granularity *G* **groups into** *H* if each granule of *H* is equal to the union of a set of granules of *G*. For example, *days* groups into *weeks*, but *weeks* do not group into *months*. A granularity *G* **is finer** than *H* if each granule of *G* is contained in one granule of *H*. For instance, *Portugal's parishes* is finer than *Portugal's districts* but *rivers* is not finer than *countries*. Some relationships are only applicable for some kind of granularities. For instance, in temporal granularities, we found *groups periodically into* or *shift equivalent* relationships. More details about granularities relationships can be found in [9], [11], [13]. Additionally, we can perform operations over granularities. In general, the operations are proposed to automate the creation of new granularities. More details about this subject can be found in [9], [11].

Camossi et al. [10] propose to represent spatio-temporal information (vector approach) in object-oriented database management systems (DBMSs) extending the ODMG standard. They define two new parametric data types. Spatial data types are defined through the $Spatial < G_s, \tau >$ data type, where G_s is a spatial granularity and τ being one of the ODMG types typically used to define conventional attributes like literal types (e.g., integer, float, etc.) or geometric types (like points, lines and polygons). Temporal or spatio-temporal data types are defined using the $Temporal < G_t, \gamma >$ data type where G_t is a temporal granularity and γ can be any mentioned data type (including a spatial data type).

To $Spatial < G_s, \tau >$ and $Temporal < G_t, \gamma >$ data types, coarse and refinement functions can be assigned allowing to hold data at multigranularities (i.e., several LoDs). Coarse functions convert data from a granularity G_α to a granularity G_β such that G_α is finer than G_β while refinement functions perform the opposite. We can have coarse or refinement functions applicable to spatial geometrical attributes or spatial quantitative and temporal attributes [10]. For example, coarse or refinement functions applied to spatial geometrical attributes can make some granules modify their position and extent, be deleted, be splitted, and be merged. Some coarse functions that can be applied on numerical types are: min, max, average. Using this approach, the user specifies, for each class attribute, what conversion functions can be used (which are already defined in [10]).

The $Spatial < G_s, \tau >$ data type index information of the type τ to spatial granules. Furthermore, the $Temporal < G_t, Spatial < G_s, \tau > >$ data type index the information of the type τ already indexed by spatial granules to temporal ones. Note that, when we define a temporal data type, the temporal granules are specifying the valid time of the information indexed to them. Another important aspect of this approach is that the indexed information will not be granules of some granularity but values of some type τ (belonging to some domain). As a result, in some scenarios, we cannot relate information at different LoDs. Consider the following class attributes: (*i*) $Spatial < G_{countries}, int >$ storing information about the exact number of population in each country; (*ii*) $Spatial < G_{countries}, String >$ also storing information about the number of population but with less precision such that the possible values are: (*i*) *less than one million* (*ii*) *one million or more and less than fifteen millions*; (*iii*) *fifteen or more millions.* Although both variables refer to the same information, we cannot relate them by stating that the former is finer than the latter. This kind of reasoning is also important to relate spatio-temporal data at different LoDs.

The proposals regarding granularities discussed so far are focused on vector data. As opposed, Pozzani et al. [13] propose a framework focused on raster data. The authors define a spatial granularity σ as a total function from two-dimensional coordinates in $\mathbb{Z}^2$ to a label set L such that $\sigma: \mathbb{Z}^2 \rightarrow L$. This way, given a cell $c \in \mathbb{Z}^2$, $\sigma(c)$ represents the label associated to c. Unlike the previous approaches, a granule corresponds to the sets of all cells sharing the same label. Then, the authors redefine the relations and operations between two spatial granularities, taking into account the definition proposed.

Either on vector-based granularities [11] or raster-based granularities [13] the authors introduce the concept of spatio-temporal granularities in order to handle with changes over time of spatial granularities. For example, a country's provinces may change over time. Spatio-temporal granularities are crucial to handle with such scenarios.

3 Granularity Theory

Granularities perform divisions of a domain. Each division corresponds to a non-decomposable entity, usually mentioned as a granule, that we can use in our statements to describe a feature of a phenomenon. As we discussed in Section 2, granularities play an important role to hold spatio-temporal data at different LoDs. However, in the literature, we just found proposals for spatial and temporal granularities. In this paper, we present a novel granularity framework in Section 3.1, allowing the definition of a granularity over any domain. Once we can define a granularity over any domain, we propose an approach that allows the transposal of the domain's properties to the granularities. This approach is fundamental to specify what properties one intends for certain granularities. Moreover, in some contexts like data mining activities, the computation of distances between granules may be needed, an issue which is also discussed. In the end, our goal is to hold spatio-temporal data at different LoDs using different granularities. To achieve such goal, we propose a granularities-based model in which every feature of a phenomenon is described by a granule. This characteristic of the model is only possible because we provide a granularity definition applicable to any domain. The model is detailed in Section 3.2.

Throughout the paper, we will use a real spatio-temporal phenomenon monitored in Portugal to illustrate our contributions. The Institute for the Conservation of Nature and Forests published detailed data on forest fires in Portugal which contains information about individual incidents. For each incident there is information about the starting point of the fire, when it started, type of burnt area(s) (agricultural land, forest, populated land), among other information.

3.1 Granularity Definition

We (as humans) are constantly using granularities, in unconscious way, in order to perform statements about phenomena. Those granularities have underlying a domain of reference. In most cases, granularities are just a way to create a domain of discourse simpler than their domains of reference. This can be observed when we use several levels of administrative divisions to make easier to refer to particular country area; it can be observed when we refer to time as days, or months; it can be perceived when we assign the age of a person always rounded to units; it can also be observed in the way we define color palettes. Here, we denote a domain of reference of a granularity as $D = (DS, RS)$ where the domain set DS corresponds to a set of elements and RS is a set of relations defined over DS. A domain set can be discrete, dense, continuous or n-dimensional. A granularity is formally defined as follows.

Definition 1 (Granularity). Let $\mathcal{IS}$ be an index set; D be a domain; 2^{DS} the power set of the DS; and GS be the set of granules extent where $GS \subseteq 2^{DS} \setminus \{\emptyset\}$ such that any two elements α and β are disjoint from each other, i.e., $\alpha \cap \beta = \emptyset$ for any $\alpha, \beta \in GS$. A granularity G is an injective and functional mapping:

$$G: GS \rightarrow \mathcal{IS} \tag{1}$$

A granularity G defines a division of a domain in set of granules. A granule g_{ind} corresponds to a pair (g, ind) where $g \in GS$ and $ind \in \mathcal{IS}$. We denote $E(g_{ind})$ as the extent of a granule which refers to g, and $I(g_{ind})$ gives the granule index which corresponds to ind. In addition, the extent of a granularity $Ext(G)$ consists in the union of the set of granules extent of G.

Unlike the majority of the proposals that can be found in the literature, we propose a mapping from the granules to an index set instead of the reverse (see Section 2.3). Using a mapping from the index set to the granules can lead too many values (from the index set) mapped to the empty set. Using our granularity definition, we just need to define the set of granules and the corresponding mapping of them to the index set.

Through this definition, it is possible to define any kind of granularity including the ones proposed in the literature (see Section 2.3). Consider the following domains examples: $D_1 = (\mathbb{R}, \{<\})$ which consist in real numbers with total order on it; $D_2 = (\mathbb{R}^2, \{<_{up}, <_{left}\})$ consisting in two-dimensional space with partial order on it. D_1 can be used to represent time while D_2 to represent space. Based on these domains, we can define the following granularities: (*i*) $S_{Districts}$ where each granule's extent corresponds to a Portugal district; (*ii*) $S_{Parishes}$ where each granule's extent corresponds to a Portugal parish; (*iii*) T_{Hours} where each granule's extent corresponds to a hour; (*iv*) T_{Days} where each granule's extent corresponds to day. Note that, in all defined granularities, the corresponding index sets are the obvious labels (e.g., in $S_{Districts}$ are the names of districts).

Additionally, to characterize the burnt area(s) for each fire incident, we can define the following discrete domain $D_3 = (\{\text{agricultural land}, \text{forest}, \text{populated land}\}, \{\})$ in order to define the following granularities: $Land_1$ where each granule corresponds to an element of D_3; and $Land_2$ with two granules $(\{\text{agricultural land}, \text{forest}\}, \text{forest areas})$ and $\{\text{populated land}\}, \text{populated land})$. These two granularities are in fact used by the data provider to describe forest fires in Portugal.

Additionally, we can define a granularity based on the extent of another granularity already defined. For instance, we can define the granularity $S_{Districts}$ based on the $Ext(S_{Parishes})$. Similarly, T_{Days} can be defined based on the $Ext(T_{Hours})$. By creating granularities based on others, we can relate granules (from different granularities) without having to look to the actual definition of them.

The granules of a granularity can be related to each other through relationships. We introduce the possibility to annotate a granularity in order to define relations between granules of a granularity. An annotation over a granularity G, denoted by $An(G)$, corresponds to a binary relation defined on the set of granules. For example, we can annotate the granularity $S_{Districts}$ with the binary relationship *adjacent*.

Definition 2 (Annotated Granularity). An annotated granularity is a granularity G and one or more annotations over it $An_1(G), \dots, An_j(G)$, denoted by $\mathcal{A}(G)$.

A granularity annotation can have specified any kind of interest relationships between granules. Some of those relations can be based on the relationships defined over the original domains, a subject that we discuss below.

Relationships between Granularities's Granules. It may exist known relationships in the domains that we would like to preserve or transpose to the granularities. For instance, a time domain is characterized by the total order. Bettini et al. [9] are actually transposing the underlying total order of the time domain to the temporal granularities through the first condition imposed to them (see Section 2.3). In the end, Bettini et al. are interested in temporal granularities where the granules are also total ordered.

Likewise, and according to the domain, we are interested to transpose the knowledge about the domain into the granularities. This way, we introduce four types of relationships which can be verified between granules of a granularity. The proposed relationships are induced from the relations held by the elements of the domain set, of the corresponding granules. The proposed types of relationships are: (*i*) complete; (*ii*) partial; (*iii*) weak; (*iv*) and, existential.

Given a granularity G defined over a domain D, a relation R defined over DS such that $R \in RS$, and g_i and g_j denotes two granules belonging to G. The formal definitions of the relationships are given, illustrated with the granules g_a and g_b belonging to a spatial granularity S specified over the domain $D = (\mathbb{R}^2, \{north\})$ such that a coordinate (x_i, y_i) is at north of a coordinate (x_j, y_j) if and only if $y_i > y_j$.

Definition 3 (Complete Relationship). A complete relationship $g_i\ R^C\ g_j$ is defined as follows.

$$g_i\ R^C g_j \Leftrightarrow \forall\ x_i \in E(g_i), x_j \in E(g_j) : x_i\ R\ x_j$$

If two granules g_i and g_j are completely related then all elements of g_i must be related with all elements of g_j through the relation R. An illustration of this relationship can be seen in Fig. 1a where g_a is completely at north of the granule g_b ($g_a\ north^C g_b$).

Definition 4 (Partial Relationship). A partial relationship $g_i\ R^P\ g_j$ is defined as follows.

$$g_i\ R^P g_j \Leftrightarrow \exists\ x_i \in E(g_i)\ \forall\ x_j \in E(g_j) : x_i\ R\ x_j\ \wedge\ \exists\ x_j \in E(g_j)\ \forall\ x_i \in E(g_i) : x_i\ R\ x_j$$

In case of two granules g_i and g_j are partially related then there is at least one element in g_i related with all elements of g_j through the relation R and similarly, there is at least one element in g_j where all elements of g_i are related with g_j through the relation R. Looking at Fig. 1b, g_a is partially at north of the granule g_b ($g_a\ north^P g_b$).

Definition 5 (Weak Relationship). A weak relationship $g_i\ R^W g_j$ is defined as follows.

$$g_i\ R^W g_j \Leftrightarrow \exists\ x_i \in E(g_i)\ \forall\ x_j \in E(g_j) : x_i\ R\ x_j \vee\ \exists\ x_j \in E(g_j)\ \forall\ x_i \in E(g_i) : x_i\ R\ x_j$$

When two granules g_i and g_j are weakly related then there is at least one element in g_i related with all the elements of g_j through the relation R or, there is at least one element in g_j where all elements of g_i are related with g_j through the relation R. Regarding the Fig. 1c, g_a is weakly at north of the granule g_b ($g_a\ north^W g_b$).

Definition 6 (Existential Relationship). An existential relationship $g_i\ R^E g_j$ is defined as follows.

$$g_i\ R^E g_j \Leftrightarrow \exists\, x_i \in E(g_i)\ \exists\, x_j \in E(g_j) : x_i\ R\ x_j$$

Finally, for two granules g_i and g_j be existentially related, it just needs that one element of each granule is related via the relation R. Considering the Fig. 1d, g_a is existentially at north of the granule g_b ($g_a\ north^E g_b$). From now on, the previous defined relationships are mentioned as the induced relations.

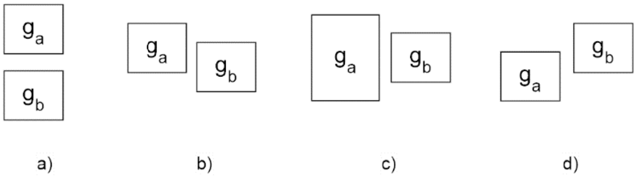

Fig. 1. Illustration of the induced relations: a) $g_a\ north^C g_b$; b)$g_a\ north^P g_b$; c) $g_a\ north^W g_b$; d) $g_a\ north^E g_b$

The induced relations are successive weakening, i.e., $g_i\ R^C g_j \Rightarrow g_i\ R^P g_j \Rightarrow g_i\ R^W g_j \Rightarrow g_i\ R^E g_j$. Therefore, given a granularity *G,* the induced relations transpose how strong a relation *R*, defined over the *DS*, is verified between two granules.

Furthermore, it is important to know what relation's properties defined over the *DS* are preserved in the induced relations. For that, we consider five properties that a relation *R* can hold: (*i*) symmetric; (*ii*) transitive; (*iii*) reflexive; (*iv*) antisymmetric; (*v*) antireflexive. It can be proved that if the relation *R* is symmetric than any induced relation is also symmetric. Furthermore, if the relation *R* is transitive then we only can state that the complete and partial relations are also transitive. For the others relations, nothing can be stated as according to the granules they may or may not inherit the transitivity property. Regarding the reflexivity property, only the existential induced relation is in any case also reflexive. Lastly, if the relation *R* is antisymmetric or antireflexive then the complete and partial relations are also antisymmetric or antireflexive, respectively. The summary of these results is displayed on Table 1.

Table 1. The induced relations properties based on the relations properties in the domain

	$g_i\ R^C g_j$	$g_i\ R^P g_j$	$g_i\ R^W g_j$	$g_i\ R^E g_j$
Symmetric	✓	✓	✓	✓
Transitive	✓	✓	inconclusive	inconclusive
Reflexive	inconclusive	inconclusive	inconclusive	✓
Antisymmetric	✓	✓	inconclusive	inconclusive
Antireflexive	✓	✓	inconclusive	inconclusive

We may use the induced relations to annotate granularities. Moreover, we can use them to specify what kind of properties we intend for certain granularities. For instance, temporal granularities defined by Bettini et al. [9] are, under our framework, granularities where the total order specified in DS induces complete relationships between their granules.

Distance Functions between Granularities' Granules. Data Mining activity plays an important role on the extraction of patterns that are hidden in very large data sets [8]. Distance/dissimilarity functions are frequently embedded into data mining approaches like clustering, classification, and nearest neighbors search. Instead of having those approaches working on the original domains, it can be advantageous if they work based on the granularities defined for such domains [12].

Suppose that there is a granularity G defined over a domain $D = (DS, RS)$, and a real-value distance function d, which quantifies the distance between elements belonging to DS such that $d: DS \times DS \rightarrow \mathbb{R}$. Additionally, g_i and g_j denote two granules belonging to G. The distances between granules can be defined based on the distances of their elements in DS. Here, we consider the following induced distances:

- **Inner Distance:** $d^I(g_i, g_j) = \min_{x_i \in E(g_i)} \min_{x_j \in E(g_j)} d(x_i, x_j)$ (2)
- **Outer Distance:** $d^O(g_i, g_j) = \max_{x_i \in E(g_i)} \max_{x_j \in E(g_j)} d(x_i, x_j)$ (3)
- **Left Distance:** $d^L(g_i, g_j) = \max_{x_i \in E(g_i)} \min_{x_j \in E(g_j)} d(x_i, x_j)$ (4)
- **Right Distance:** $d^R(g_i, g_j) = \min_{x_i \in E(g_i)} \max_{x_j \in E(g_j)} d(x_i, x_j)$ (5)

The inner distance corresponds to the minimum distance between two granules while the outer distance is the maximum distance. Moreover, the left distance corresponds to the Hausdorff distance from g_i to g_j while the right distance corresponds to the Hausdorff distance from g_j to g_i. Besides the induced distances introduced, it can be defined several other distances like the distance between centers of mass of the granules, the minimum between the inner and the outer distance, and so on.

Relationships between Granularities. Recall that, the relationships between granularities allow one to compare and relate granules belonging to different granularities, useful to hold spatio-temporal data at different LoDs. In this section, we revisit the majority of the relationships introduced in the literature considering the proposed granularity definition.

Two granularities G and H can be related as follows. For a matter of simplification, in the following formal definitions, we refer to a granule of a granularity by using the lower case letter of the corresponding letter of the granularity. For instance, *each granule's extent of h* can be stated as *each h's extent*.

- $\boldsymbol{G}$ **groups into** $\boldsymbol{H}$ **(**$\boldsymbol{G \trianglelefteq H}$**):** each h's extent is equal to the union of a set of g's extent. The formal definition is: $\forall h \in H, \exists G' \subseteq G: \bigcup_{g' \in G'} E(g') = \mathrm{E}(h)$
- $\boldsymbol{G}$ **is finer than** $\boldsymbol{H}$ **(**$\boldsymbol{G \preccurlyeq H}$**):** each g's extent is contained in one h's extent. The formal definition is: $\forall g \in G, \exists h \in H: E(g) \subseteq E(h)$
- $\boldsymbol{G}$ **is a sub-granularity** $\boldsymbol{H}$ **(**$\boldsymbol{G \sqsubseteq H}$**):** for each g there is one h with the same extent. The formal definition is: $\forall g \in G, \exists h \in H: E(g) = E(h)$
- $\boldsymbol{G}$ **is equivalent to** $\boldsymbol{H}$ **(**$\boldsymbol{G \asymp H}$**) :** $G \sqsubseteq H$ and $H \sqsubseteq G$
- $\boldsymbol{G}$ **partitions** $\boldsymbol{H}$ **(**$\boldsymbol{G \oplus H}$**):** $G \trianglelefteq H$ and $G \preccurlyeq H$
- $\boldsymbol{G}$ **is extent covered by** $\boldsymbol{\mathcal{H}}$ **(**$\boldsymbol{G \,\hat{\sqsubseteq}\, H}$**):** the extent of G is contained in the extent of H, formally defined as: $Ext(G) \subseteq Ext(H)$

Some properties of the set of relationships revisited are: $G \sqsubseteq H$ implies $G \preccurlyeq H$, which in turn implies $G \,\hat{\sqsubseteq}\, H$. Furthermore, $G \preccurlyeq H$ does not implies $G \trianglelefteq H$, nor the vice-versa.

The relationships groups into, finer than, sub-granularity, partitions, and the extent covered (equivalent to image covered in the literature) are redefinitions of the relationships proposed and used in other works [9], [11] (see Section 2.3).

Through the equivalent relationship, we intend a relationship capable to relate different granularities containing granules with equal extent. For example, we can have two spatial granularities where each granule corresponds to a country of our world. One granularity can be indexing the granules using native names and the other English names. Moreover, the granularities, related to the example about forest fires in Portugal, are related in the following manner: (*i*) $S_{Parishes}$ partitions $S_{District}$; (*ii*) T_{Hours} partitions S_{Days}; (*iii*) $Land_1$ is finer than $Land_2$.

Spatio-temporal Granularities and Granularities Evolution. Granularities may change over time. For instance, countries' provinces may change over time i.e. some provinces can be merged, other can be splitted or new provinces can appeared; in biology, kingdom is a classification of life used to rank organisms which is a classification that have been suffering changes over time. Thus, if we define the *countries' provinces* or *kingdom* granularity, we need to be able to handle with the evolution of the corresponding granularities over time.

In the literature, regarding just spatial granularities, this issue is handled under the terminology of spatio-temporal granularities, which corresponds to an evolution of a spatial granularity over time, from our point of view.

Through Definition 1, we allow the definition of a granularity over any domain. Consequently, we need to handle with an evolution of granularity that can be defined over any domain, something that we left for future work. However, we envisage addressing this issue under a different terminology. Instead of using the term spatio-temporal granularity, we propose to use the term evolution of a granularity. This way, we reserve the term spatio-temporal granularity to mention granularities where each granule refers to a portion of a $\mathbb{R}^3$ (e.g, if we assume the space as $\mathbb{R}^2$ and time as an additional dimension).

3.2 Granularities-Based Model

The granularities play a key role in the level of detail that we perform statements about the reality. In general, for any domain, but in particular to time and space domains, several granularities are needed in order to allow statements at different LoDs. For example, in the case of forest fires in Portugal, one statement may describe the occurrence of an incident whose origin start at a particular hour (a granule belonging to the granularity T_{Hours}) and, it takes place in a particular parish (which corresponds to a granule of $S_{Parishes}$). Others statements, involving the type of burnt area, can be performed using the granules of granularities T_{Days}, $S_{Parishes}$, $Land_1$. In short, the granularities are the instruments that allow modeling the different LoDs acceptable in the statements which are performed about phenomena.

In order to reason at different LoDs about spatio-temporal natural or human activities a granularities-based model is proposed, formally defined as follows.

Definition 7 (Granularities-based Model). Let $\mathcal{G} = \{\mathcal{A}(A), \ldots, \mathcal{A}(Z)\}$ be the set of annotated granularities, and $\mathcal{P}$ a set of predicates. A model $\mathcal{M}$consists in a set of atoms where the arguments are granules of granularities in $\mathcal{G}$.

An atom is a predicate symbol together with their arguments. In a granularities-based model the predicates' arguments are granules belonging to granularities in $\mathcal{G}$. Thereby, atoms are used to describe phenomena. A fundamental characteristic that makes our approach different from others is the fact that every feature of a phenomenon is described by granule that refers to a granularity. Any granule of one granularity can be related to other one from a different granularity. As a result, different atoms, forming a model, can be related in multiple ways. Particularly, to reason and relate atoms at different LoDs, it is crucial that the granularities of a model, regarding the same domain, are related through *groups into*, *finer-than* or, *partitions* relationships.

Consider again data about forest fires in Portugal. For this phenomenon, there were already defined two spatial granularities, two temporal granularities, and two granularities regarding the burnt area type. To illustrate the proposed model, we introduce three more granularities: S_{coords}, Ar_1, Ar_2. The granularity S_{coords} is defined over D_2 (introduced in Section 3.1) whose granules are a subset of the elements of the domain such that S_{coords} is finer than $S_{Parishes}$. Furthermore, the granularities Ar_1 and Ar_2 allow to make statements involving the size of the burnt area (in hectares). This way, the Ar_1 was defined over real numbers domain and its granules correspond to ([1, 250[, tiny area), ([250, 1000[, medium area), ([1000, 2000[, big area), and ([2000, 5000], huge area); the Ar_2 was defined with the following granules ([1, 10000[, small area), ([10000, 80000], large area). Note that, Ar_1 is finer than Ar_2. To make statements about the previous phenomenon, we can use the following predicate: *incident (where, when, type of land, burnt area)*. When we apply this predicate with the appropriate granules, we are describing that an incident occurred in a particular place, at particular time and with a specific burnt area (in hectares) of a type of land.

Based on the defined granularities, we are in conditions to define a model to describe forest fires in Portugal. The granularities provide the domain of discourse that we can use in a model to perform statements. Hence, we present a few illustrative atoms of a possible model:

Example 1. incident $(\mathrm{s}_{\mathrm{coords1}}, \mathrm{t}_{04-08-\ 2006\ 17\mathrm{h}}, \mathrm{land1}_{\mathrm{forest}}, \mathrm{ar1}_{\mathrm{medium\ area}})$
Example 2. incident $(\mathrm{s}_{\mathrm{coords2}}, \mathrm{t}_{25-02-\ 2001}, \mathrm{land2}_{\mathrm{populated\ land}}, \mathrm{ar1}_{\mathrm{tiny\ area}})$

In the atom one, it is described that a fire incident occurred in a particular latitude and longitude, starting at 4th August 2006, at 17:00 hours burning a medium forest area; in the atom two, it's stated that a fire incident occurred in a specific latitude and longitude at 25th February 2001, burning a tiny populated area. As can be observed, in a granularities-based model, the statements can be performed using different granularities i.e., with different levels of detail.

Moreover, and more relevant to our work, the granularities-based model provides the ability to perform statements that summarizes what happened. To illustrate such ability, we introduce the following predicates: (*i*) *incidents(where, number of incidents, burnt area)* which allows describing the number of incidents and the corresponding burnt area in a spatial granule; (*ii*) *incidents(when, number of incidents, burnt area)* allows to state the number of incidents and the corresponding burnt area in a temporal granule; (*iii*) *incidents(where, when, number of incidents, burnt area)* which permits to describe the number of incidents occurred, and the respectively burnt area, in a given spatial and temporal granule. In order to describe the number of incidents, we introduce a new granularity C defined over the natural numbers domain, and its granules consist in the elements of the domain. Based on the previous predicates, we display the following atoms[1]:

Example 3. incidents $(\mathrm{s}_{\mathrm{Mação}}, \mathrm{c}_{80}, \mathrm{ar2}_{\mathrm{small\ area}})$
Example 4. incidents $(\mathrm{t}_{1-08-\ 2003}, \mathrm{c}_{493}, \mathrm{ar2}_{\mathrm{large\ area}})$
Example 5. incidents $(\mathrm{s}_{\mathrm{Porto}}, \mathrm{t}_{23-06-\ 2003\ 23\mathrm{h}}, \mathrm{c}_{44}, \mathrm{ar2}_{\mathrm{small\ area}})$

In the atom three, it's stated that in the parish of *Mação* occurred 80 incidents which resulted in a small burnt area; in the atom four, it's described that at 1st August 2003 occurred 493 fire incidents that result in a large burnt area; regarding atom five, it's stated that 44 fire incidents occurred, in district of *Porto*, at 23th June 2003, at 23:00 hours which result in a small burnt area. Note that, in the previous examples, it were performed the appropriate aggregation operations, namely, count for counting the number of fire incidents and sum to acquire the total amount of the burnt area.

These atoms provide high level descriptions about fire incidents which occurred in a spatial granule, temporal granule, or both, accordingly. The ability to reduce the amount of data through high level descriptions can make easier the achievement of data insight. For instance, the predicate *incidents(where, number of incidents, burnt area)* can be advantageous to identify what regions are more affected by forest fires, instead of analyzing information provided by atoms referring individual fire incidences. Therefore, the loss of precision provided by the predicates *incidents* can be suitable for certain analytical activities that otherwise would be challenging.

[1] *Porto* is a Portugal district while *Mação* is a parish.

Once predicates arguments are granules of granularities in the model, we have the ability to relate different atoms via relationship between granules. Consider that, the granules $s_{coords1}$ and $s_{coords2}$ are within the granule $s_{Mação}$. As a result, we can relate the atom three with the atoms one and two.

The previous illustrative model shows an example regarding spatio-temporal events. However, different spatio-temporal phenomenon behave in different manners [14]. In addition to spatio-temporal events, we can have, for instance, georeferenced time-series or moving objects/entities. The georeferenced time-series refers to entities or objects that have a spatial fixed location over time, recording numerical information (e.g., sales in stores, temperatures recorded by meteorological stations). Moving objects/entities are spatial objects/entities that change their location over time (e.g., movement of persons, cars, boats, among others). Through the proposed model, different spatio-temporal phenomenon can be handled, expressing them at different LoDs.

4 Conclusions and Future Work

Our goal is to provide an approach to deal and relate different LoDs about spatio-temporal phenomena. To achieve that goal, we analyze the literature proposals for dealing with spatio-temporal phenomena at different levels of detail. Throughout this process, the concept of granularity drew our attention as well as the ability to relate different granularities. We propose a framework that generalizes the granularity definition to any domain, covering the definitions of temporal and spatial granularities proposed in the literature. Furthermore, we propose an approach to transpose knowledge about the relations between elements of the original domains and its properties to the granularities[2]. This allows to annotate granularities with the induced relationships.

Based on the granularity framework, it is presented a granularities-based model to deal and relate different LoDs about spatio-temporal data. Unlike other works, our model describes every feature of a phenomenon through a granule that belongs to a defined granularity. This allows relating atoms at different granularities in multiple ways through the relationships between granularities.

As future work we want to extend the definition of granularity in order to handle the evolution of granularities over time, and also introduce the concept of spatio-temporal granularities, as we briefly discussed in Section 2.3 and Section 3.1. Furthermore, we want to deeply study in what conditions an induced distance can inherit properties of the distance defined on the domain set. Regarding the model, future work can be directed for the characterization of types of models defined for particular kinds of data; moreover, for each models' types we are interested to develop automatic operations to generate predicates with syntheses about the reality; furthermore,

[2] The proofs regarding the results presented in the Table 1 can be found in: `http://staresearch.net/resources/Papers/2014/2014-ICCSA-GranularitiesDemonstrations.pdf`

given a model and an analytical goal, we are interested in identifying what LoDs are appropriate for the analytical task.

References

1. Keim, D.A., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual Analytics: Definition, Process, and Challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008)
2. Zadeh, L.A.: Fuzzy sets. Information and control 8(3), 338–353 (1965)
3. Parent, C., Spaccapietra, S., Vangenot, C., Zimányi, E.: Multiple Representation Modeling. In: Liu, L., Özsu, M.T. (eds.) Encyclopedia of Database Systems, pp. 1844–1849. Springer US (1849)
4. Parent, C., Spaccapietra, S., Zimányi, E.: The MurMur project: Modeling and querying multi-representation spatio-temporal databases. Information Systems 31(8), 733–769 (2006)
5. Zhou, X., Prasher, S., Sun, S., Xu, K.: Multiresolution Spatial Databases: Making Web-Based Spatial Applications Faster. In: Yu, J.X., Lin, X., Lu, H., Zhang, Y. (eds.) APWeb 2004. LNCS, vol. 3007, pp. 36–47. Springer, Heidelberg (2004)
6. Weibel, R., Dutton, G.: Generalising spatial data and dealing with multiple representations. Geographical Information Systems 1, 125–155 (1999)
7. Stell, J., Worboys, M.: Stratified Map Spaces: A Formal Basis for Multi-resolution Spatial Databases. In: Proceedings 8th International Symposium on Spatial Data Handling, pp. 180–189 (1998)
8. Committee on the Analysis of Massive Data, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council: Frontiers in Massive Data Analysis. National Academies Press (2013)
9. Bettini, C., Jajodia, S., Wang, S.: Time Granularities in Databases, Data Mining, and Temporal Reasoning. Springer (2000)
10. Camossi, E., Bertolotto, M., Bertino, E.: A multigranular object-oriented framework supporting spatio- temporal granularity conversions. International Journal of Geographical Information Science 20(5), 511–534 (2006)
11. Belussi, A., Combi, C., Pozzani, G.: Formal and conceptual modeling of spatio-temporal granularities. In: Proceedings of the 2009 International Database Engineering & Applications Symposium on IDEAS 2009, p. 275 (2009)
12. Camossi, E., Bertolotto, M., Kechadi, T.: Mining Spatio-Temporal Data at Different Levels of Detail. In: The European Information Society, pp. 225–240. Springer (2008)
13. Pozzani, G., Zimányi, E.: Defining spatio-temporal granularities for raster data. In: MacKinnon, L.M. (ed.) BNCOD 2010. LNCS, vol. 6121, pp. 96–107. Springer, Heidelberg (2012)
14. Kisilevich, S., Mansmann, F., Nanni, M., Rinizivillo, S.: Spatio-Temporal clustering: a Survey. ISTI-CNR, Tech. rep. (2010)