Towards Intelligent Analysis of Complex Networks in Spatial Data Warehouses

Sergio Izack Vieira Filho\textsuperscript{1,4}, Renato Fileto\textsuperscript{1,2},
Andre Salvaro Furtado\textsuperscript{2,5}, Ricardo Haus Guembarovski\textsuperscript{3,4}

\textsuperscript{1}Post-Graduate Program in Computer Science (PPGCC)
\textsuperscript{2}Informatics and Statistics Department (INE)
\textsuperscript{3}Post-Graduate Program in Knowledge Engineering (EGC)
Federal University of Santa Catarina (UFSC)
PO BOX 476, 88040-900 - Florianópolis-SC, BRAZIL

\textsuperscript{4}Santa Catarina Electrical Central (CELESC)
Av Itamarati, 160 - 88034-900 - Florianópolis-SC, BRAZIL

\textsuperscript{5}Geography Department (DG), Santa Catarina State University (UDESC)
Av. Madre Benvenuta, 2007 - Itacorubi - 88.035-001 - Florianópolis-SC, BRAZIL

\{sergiolvf, fileto, asalvaro\}@inf.ufsc.br, ricardohg@celesc.com.br

Abstract. This paper proposes a knowledge-based model to support spatial and temporal analysis of complex networks in data warehouses. The proposed model, inspired from geography, represents the dynamics of spatial elements in the network. Ontologies, describing element classes from a domain, space partitions, and time periods, respectively, help to define dimensions of data marts customized for specific analysis needs. This model has been implemented in a prototypical framework, with a GUI supporting spatial OLAP and visualization of the varying state of network portions in graphs and maps. It supports the investigation of spatial and temporal patterns and tendencies, as has been observed in an electricity distribution case study.

1. Introduction

On-Line Analytical Processing (OLAP) [Kimball and Ross 2002] in traditional data warehouses (DW) organized according to the conventional multidimensional model can only handle categoric dimensions and numeric measures [Inmon 2005]. On the other hand, some studies show that around 80\% of data are related to some spatial information [Malinowski and Zimányi 2008]. GIS [Chang 2007] are also important to provide information for decision making purposes. The composition of these technologies, DW and GIS, can be useful for several applications, such as the spatial analysis of bird migration, traffic behavior, and power grids [Malinowski and Zimányi 2008]. Spatial data warehouses (SDW) [Bimonte et al. 2006, Escribano et al. 2007] try to reconcile these two successful technologies. In SDW’s multi-dimensional and spatial data models [Damiani and Spaccapietra 2006, Malinowski and Zimányi 2008], spatial objects can arise as members of dimensions (e.g., polygons representing states and cities) and as measures of fact tables (e.g., geographic points where equipments are installed). Spatial On-Line Analytical Processing (SOLAP) [Rivest et al. 2005] handle spatial elements integrated to the dimensional model, using a wide variety of operators and aggregation functions for handling spatial objects [Ruiz and Times 2009], besides the traditional OLAP...
operators and aggregation functions used to handle conventional data. It enables information analysis using geographic features, where topological relations are used to filter and aggregate conventional and spatial measures [Malinowski and Zimányi 2008].

However, current SDW technologies have limited means to handle complex networks of spatial elements (objects located in the space). Complex networks have been used to model and analyze a wide variety of systems and phenomena, ranging from the Internet and the social Web to traffic and power grids [Strogatz 2001, Newman 2003]. This paper proposes a model for handling complex network data in a SDW. This model represents spatial elements, their functions, connections and dynamics as proposed by [Santos 2008]. The forms of these elements are represented as geographic extensions. Their functions are defined by concepts from domain ontologies. These concepts also influence the measures that can be taken from the spatial elements taking part in a process. The domain ontology, along with an ontology about spatial partitions and another ontology about time periods, help to define the dimensions of data marts. We have built a framework based on this model which has a knowledge-based graphical user interface (GUI) for domain experts with basic TI skills to use SOLAP to analyze large amounts of data from vast complex networks, and visualize the results in graphs and maps. The facilities of this GUI include sliders for the user to track the temporal changes of measures of the state of spatial elements in the complex network (such as the load on energy distribution equipments), according to absolute and cyclic time periods. These facilities have been tested using a case study in the area of electricity distribution.

1.1. Related Work

Several SDW models have been proposed in the literature [Damiani and Spaccapietra 2006, Malinowski and Zimányi 2008]. In addition to conventional and spatial data, many applications involve the handling of historical series. Representations of time have to take into account the granularity of the periods of time and cycles of time [Malinowski and Zimányi 2008]. Thus, some multidimensional models have been enriched with temporal [Mendelzon and Vaisman 2000, Moreno et al. 2009] and spatio-temporal [Bertino et al. 2009, Golfaelli and Rizzi 2009] features, giving rise to spatio-temporal data warehouses (STDW) [Savary et al. 2004]. However these models do not consider spatial elements connected in a network.

Works about visualization of the state of complex networks, such as those proposed for the analysis of power grids [Overbye et al. 2007, Moreno-Munoz et al. 2009], on the other hand, do not provide systematic means like SOLAP for the analysis of large volumes of spatial and temporal data from these networks. Other works address the analysis and visualization of trajectories, in applications like the analysis of the traffic [Braz et al. 2007, Leonardi et al. 2010], but they do not incorporate the topology of the spatial network and precise measures of the state of network elements in the data model.

Dynamic networks of elements on earth have been addressed by geography [Santos 2006, Santos 2008, Cámara et al. 2001]. In order to give a better understanding of the complex relationships of geographical elements over time [Santos 2008] proposes to model such elements with respect to four aspects: form, function, structure (connections), and the processes that the connected elements participate. [Cámara et al. 2001] suggests that future geographic information systems would use such ideas from the geog-
raphy, along with means for knowledge representation, such as ontologies, to cope with dynamic systems in the space.

The remainder of this paper starts with a review of some foundations from geography in Section 2. Section 3 describes the proposed model for handling data from complex networks of spatial elements in data warehouses. Section 4 presents some information analysis using the proposed knowledge-based model and GUI in a case study. Finally, Section 5 enumerates the conclusions and future work.

2. The Geographic Space

The geographic space has such a complexity that it requires an analytical model that fragments the totality of the objects and phenomena taking place in the space [Santos 2006]. A model for spatial analysis of networks of spatial elements connected or interacting with each other is proposed [Santos 2008]. This model is based on descriptions of complex spatial elements, their interrelations, and measures characterizing their state, according to four aspects: form, function, structure, and process.

The form is the visible aspect of a spatial element in a given instant. It can be influenced, though in an imperfect way, by the role of the spatial element in an information analysis from a particular viewpoint. The form carries an idea of the purpose and relevance of the spatial element, for performing one or more functions in a system.

The function is the action performed by or with the use of the spatial element. It is what is expected to be executed by the spatial element with a given form. The same form of a spatial element can have one or more functions, according to the context, which includes the space and time where the element is inserted. This variation is the consequence of constant changes in the physical environment and the society. Though some forms can last for long periods, they can have new functions assigned to them.

The relationships between spatial elements, i.e., the components of the spatial whole, results in the structure. The form-function pairs of the spatial elements can be organized according to flows of relationships among elementary units and their significant combinations [Perroux 1969].

The process, by its turn, is a continuous action developing towards some result, which can only be apprehended in a spatio-temporal view [Santos 2008]. By analyzing the process it can be possible to understand the patterns and trends in the evolution of the structure, and the interactions among spatial elements to reach a given result.

The conceptualization of these aspects of the description of spatial elements and their connections is necessary for a complete description of what happens in networks of elements in the space as time goes on. It allows a concrete and precise interpretation of the evolving processes in the geographical space [Santos 2008].

In the real world, the concepts for describing spatial elements are highly dependent on the application domain. For example, the evolution of the traffic can be analyzed in a mesh composed of highways, roads, streets, and other elements of transport networks. On the other hand electricity distribution and consumption occur in networks composed of power generation stations, wires, switches, voltage converters, and consumption meters, among other equipments. However, the partitions of the space and the time periods used
for information analysis have many concepts that apply to a variety of domains. Thus, the
time and space conceptualizations necessary for information analysis can sometimes be
reused in different domains with some customization.

3. The Proposed Model

Our dimensional model for handling information from complex networks in spatial data
warehouses extends the model proposed in [Malinowski and Zimányi 2008]. It also relies
on ideas of [Santos 2008] and ontologies. Due to space limitations we abstract details of
the dimensional model in this paper. We show how our model organizes information of a
complex network of spatial elements and how it supports analysis of this information.

3.1. Complex Networks of Spatial Elements

A complex network in our model is represented as a directed graph \( G(V, E) \). Each vertex
\( v \in V \) represents a spatial element. Each directed edge \( e \in E \) represents a connection
from an element \( v \) to an element \( v' \) (\( v, v' \in V \)). Each vertex \( v \in V \) has the form:

\[
v = (\Gamma, \Phi, P(v, t), N(v, t), S(v, t))
\]

where:

- \( \Gamma \) is a set of forms of \( v \) (including spatial location in some coordinate system)\(^1\),
- \( \Phi \) is a set of functions (roles) of \( v \) (functions can be defined by concepts in the
domain ontology and be associated with forms),
- \( P(v, t) : V \times T \rightarrow 2^{(V-v)} \) is a function that gives the preceding elements of \( v \) in
the complex network in a given time period \( t \in T \),
- \( N(v, t) : V \times T \rightarrow 2^{(V-v)} \) is a function that gives the following elements of \( v \) in
the complex network in a given time period \( t \in T \), and
- \( S(v, t) : V \times T \rightarrow \mathbb{R}^n, (n \in \mathbb{N}, n > 0) \) is a function that gives the values of a set
of \( n \) state variables of the spatial element \( v \) (each spatial element can be a complex
non-linear system with \( n \) state variables [Strogatz 2001]).

The time \( T \) is represented a sequence of periods of some duration (e.g., hour). The
complex network \( G \) has a connection \( e = (v, v') \in E \) from vertex \( v \in V \) to vertex \( v' \in V \)
in a given time period \( t \in T \) if and only if \( v' \in N(v, t) \) and \( v \in P(v', t) \). For simplicity we
only consider in this paper the dynamics of the internal states \( S(v, t) \) varying in time) for
each \( v \in V \), but not the dynamics of the network topology. In other words, we consider
the set of network links \( E \) fixed, i.e., \( \forall v \in V, t, t' \in T : P(v, t) = P(v, t') \land N(v, t) = N(v, t') \).

3.2. Linked Spatial Elements

Figure 1 illustrates the Complex GeoObjects representing spatial elements of complex net-
works in our model. The methods for accessing these objects cover the four aspects of the
geographic reality (form, function, structure and process) introduced by [Santos 2008].
The method \( \text{getForm} \) returns the form(s) of the object valid for the context provided as
argument. The context can include a geographic region (defined by a concept from the
Space Partitions Ontology), a time period (defined by a concept from the Time Periods
Ontology) and a scale. The set of returned forms varies with the context and the state of

\(^1\) We intentionally avoid commitment to any standard for representing forms.
the object in the time period specified in the context. The method \textit{getFunction} returns one or more concepts from the domain ontology describing the function(s) of the object in the context provided as argument. The methods \textit{previousObjects} and \textit{nextObjects} return the sets of objects representing previous and next elements in the complex network, respectively. With these methods one can trace the structure of connected components of the complex network. Finally, \textit{getValue}, returns the value of an aggregate measure (e.g., maximum, minimum, average) of a variable of the internal state of the spatial object in the context provided as argument. It enables the analysis of the processes occurring in the complex network, by giving access to information about the time varying state of each spatial element in the complex network.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Objects representing spatial elements of complex networks}
\end{figure}

3.3. Ontologies

The three ontologies referenced in Figure 1 describe the spatial elements and support consolidated analysis of the state of different parts of the complex network along time. Figures 2-a, 2-b and 3 present the top conceptualizations of these ontologies\textsuperscript{2}. Analytical dimensions of data marts for addressing specific analysis needs can be specified as views of these ontologies, i.e., extracts with selected concepts, instances and relationships. These subsets constrain the data mart to some focus, avoiding a large, cumbersome and inefficient analytical cube.

The \textit{Spatial Partitions Ontology} and the \textit{Time Periods Ontology} are more stable in our model than the \textit{Domain Ontology}, which describes and classifies specific domain

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Fragments of ontologies used in the proposed model}
\end{figure}

\textsuperscript{2}These are just illustrative portions of the complete ontologies. Many concepts, semantic relationships and all instances are omitted for simplicity and due to space limitation.
concepts. Thus, the latter needs to be changed in order to support information analysis for different application domains. The major concepts of the former ontologies, require minor customizations for different domain. For example, the analysis of electric power grids can use different regions and time periods (more related to social life) than those used for the analysis of and environmental issues (more related to nature).

3.3.1. Spatial Partitions Ontology

Analysis of spatial information requires spatial knowledge, including:

- relationships between spatial partitions and between components of these partitions (e.g., states, which composed of counties);
- homonyms, i.e., different entities referenced by the same name (e.g., Santa Catarina referring to a Brazilian State or the island with the same name in that state);
- synonyms, i.e., alternative names referring to the same entity (e.g., Florianópolis, Floripa or The Island of Magic referring to the capital of Santa Catarina State).

The Spatial Partitions Ontology addresses all these issues. It describes hierarchies of partitions of the space in land parcels that can be relevant for information analysis. Land parcels can be countries, states, cities and other regions defined by some criteria (e.g., relief, vegetation, economic and demographic issues). Figure 2-a shows that both State and City are specializations of LandParcel, and that an instance of State is composed of instances of City or, alternatively, by instances of Regional, another specialization of LandParcel to divide the territory for energy distribution purposes. The Spatial Partitions Ontology also includes instances of these concepts, containment relationships between instances (e.g., Santa Catarina, the state, contains the city of Florianópolis) and synonyms. Thus, it can be used as a gazetteer to identify and help to solve ambiguities.

3.3.2. Time Periods Ontology

Analysis of a complex network dynamics can refer to two major kinds of time intervals:

- linear time periods, which succeed each other in the calendar, such as specific days, weeks, months and years;
- cyclic time periods, which repeat in cycles, such as daily periods (e.g., morning, afternoon, evening and night) and yearly periods (e.g., the year seasons).

The Time Periods Ontology classifies and describes different kinds of time periods used for information analysis. One user can be interested in analyzing the changing state of portions of a complex network according to linear time (e.g., the evolution of a measure as time goes on) or some cyclic time periods (e.g., the variation pattern of some measure along the seasons or daily periods, considering all the recorded years or days). Figure 2-b shows an extract of the ontology of time periods. This ontology addresses, among other details, the classification on cyclic and absolute time periods, their relationships, and spatial customization of certain time periods according to the geographic location (e.g., the seasons of the year occur in different months in the northern and southern hemispheres).
3.3.3. Electricity Distribution Equipment Ontology

The Domain Ontology describes the hierarchy of classes of spatial elements. Concepts of this ontology are used in our model to define the functions of the spatial elements, according to domain specific knowledge. Figure 3 illustrates the high level class hierarchy on domain ontology about components of an electrical distribution grid, used in our case study. This ontology classifies the electricity distribution equipments according to their function and voltage range.

![Figure 3. Electrical grid equipments ontology](image)

4. Case Study

The electricity distribution companies provide services for different categories of consumers: residential, business, industrial and rural. The Brazilian Electricity Regulatory Agency (ANEEL) imposes certain rules relative to the quality levels that must be observed for these services. These rules force the companies to develop solutions to optimize their operation and reach quality standards. The control of their activities, the planning of investments and the understanding of technical problems and limitations require the spatial and temporal analysis of the performance of their network of electricity distribution assets. The domain expert needs systematic ways to assess the temporal evolution of the topology and the operation (level of load, number of faults) of the electricity network equipments installed in the geographic space. The analysis of the behavior of these variables is fundamental for right and fast decision making.

4.1. Data Mart for Analyzing an Electricity Distribution Grid

Figure 4 presents the star schema of a data mart developed for the load analysis of the power grid of an electricity distributor: Santa Catarina Electrical Central (CELESC). The three analytical dimensions of this data mart, Space, Time, and Equipment are views of the ontologies presented in section 3.3 specified for addressing some specific analysis needs in this data mart. The levels of the dimension Space are specializations of the concept Land Parcel: State, Regional (set of cities from a state under the same electricity distribution manager), City and UrbanArea. The members of these levels have geographical extents to represent the respective land parcels in maps. Some members are associated to each other according to containment relationships (e.g., Florianópolis is a city of Santa Catarina State). The level equipmentType of the dimensional Equipment has as members the types of electricity distribution equipments described in the domain ontology of
These equipments can be owned by CELESC or third parties, including consumers in some cases. The fact table maintains the numeric measure \textit{LoadPercentage} (the maximum, minimum or average load percentage for all the equipments installed in a certain region during a time period) and the spatial measure \textit{Equipments} (complex objects representing the equipments and measures of their state along time, as described in section 3).

### 4.2. Information Analysis

The model proposed in section 3 supports analysis of information from complex networks on the geographic space using OLAP and tracing of the state of the interconnected components of the network in different time periods. In the following we present some spatial information analysis that can be done on this model, using the data mart illustrated in figure 4 as a case study.

#### 4.2.1. Spatial OLAP

The user can initiate his analysis of an extensive complex network by using SOLAP to have an overview of the situation and investigate the consolidated state of the network in different regions and time periods. SOLAP provide systematic means for doing information analysis for strategic and tactic purposes, before going into operational details of the network. Figure 5 illustrates the results of a sequence of \textit{drill-down} operations applied to the data mart illustrated in figure 4 to investigate the distribution of the maximum load percentage of the capacity of the power grid across different regions. As the user identifies the regions with high load he can \textit{drill-down} into these regions to see zoomed-in maps of the distribution of the load in their sub regions. Figure 5(a) shows the distribution of the maximum load percentage of electricity distribution equipments across the cities in the Florianópolis metropolitan area, Figure 5(b) shows details of the load in Tijucas, a particularly loaded city, and Figure 5(c) shows the load on specific equipments installed its urban area. These maximum loads are taken from the SDW measures \( S(v, t) \) of each spatial element \( v \) contained in the corresponding region, with \( t = \text{April } 2010 \).

The calculus of the measure \textit{LoadPercentage} presented in figure 5 for each spatial element is based on the relation between the nominal power (kVA) and the energy demand. The nominal power is obtained by summing the power supplied by the medium to
low voltage transformers located in the each land parcel. The demand is derived from the total energy consumption (kWh) measured in the consumption units. This estimation can be done by using a statistic approximation function. The calculus of the measures follows a sequence of steps for each land parcel\textsuperscript{3}:

1. search for the lvsegments connected in each transformers and consumption units located in the land parcel;
2. sum the nominal power of the transformers in the parcel;
3. sum the power consumed by the consumption units in the parcel, and derive the maximum demand;
4. calculate the relation between the nominal power and the maximum demand.

\textbf{4.2.2. Trace Analysis of the Network}

The trace operation analyses the state of the connected spatial elements of portions of the complex network in different time periods. In the case study of the electric energy distribution it is possible to assess, for example, the energy flow across the equipments installed in the geographic space. Figure 6 shows the evolution of the maximum load of a major electrical energy feeder located in Florianópolis downtown. The user can move the time slides under the graphic showing the variation of the overall load on the feeder in different months (in the right side) in order to visualize the temporal evolution of the distribution of the load in the map (in the left side), month by month. Figure 6(a) shows the load in May 2009 and Figure 6(b) in December 2009. These graphs plotted by following the function \( N(v, t) \) (next spatial element) for each equipment \( v \), starting in the \textit{Feeder Start Point} for the respective month.

The maps presented in Figure 6 show an estimation of the maximum electrical current flowing on each distribution equipment. This measure is calculated from an estimation of the maximum demand on each equipment, using the formula:

\[
I = \frac{DEM}{V \times \sqrt{3}}
\]

\textbf{Where:}

\textsuperscript{3}We consider the time period fix in order to simplify the explanations. The measures can be aggregated for different time periods as well.
Figure 6. Spatio-temporal analysis of the load in a power grid portion

$I$ is the maximum current consumed (A);

$DEM$ is the maximum estimated demand (kVA); and

$V$ is the operation voltage (V).

The estimation of the demand on the equipments begins in the feeder start point, where it is estimated using the total current injected in the circuit and the total energy consumption. The demand for each other equipment in the same feeder is estimated using the demand in the previous equipments (previous elements in the complex network) and the consumption in the connected consumption units (leaves of the transitive closure of following elements in the network).

Finally, some issues observed in Figure 6 deserve explanations:

- Outliers are due to error in the collection of basic measures such as the energy consumption (from the consumer’s meters) and commercial losses due to factors like the joule effect, and energy thieving.
- Some measures of energy consumption are higher than the energy injected in the corresponding portion of the circuit, due to measurement faults and, sometimes, due to modification in the topology of the power grid which are not communicated to electricity supply company.
- The temporal analysis of the load enables the identification of some seasonal patterns. The graphic in Figure 6 shows that the energy consumption is higher during the summer (due to the high proportion of tourists in Florianópolis at this time and the use of air conditioning) and that there has been a vegetative growth in the energy consumption during the last two years.
5. Conclusions and Future Work

The knowledge-based model presented in this paper enables the representation and analysis of historical series of data about the state of interconnected elements of complex networks in a spatial data warehouse. The measures about the state of the network components are estimated and aggregated according to the network topology and a dimensional schema. It supports the analysis of portions of the network using SOLAP, as well as tracing the state of network portions. These analyses are oriented by analytical dimensions like element types, space, and time, which can be derived from ontologies describing the respective conceptualizations. Information visualization in graphs and maps facilitates the identification of spatial and temporal patterns and tendencies.

Currently the proposed model is being tested using a case study in the area of electricity distribution. Future work include: (i) improve the representation of the network topology and dynamic states of the network elements, in order to efficiently store data and process queries for information analysis; (ii) use data sampling from sensors installed in different kinds of network elements to calibrate methods to estimate the distribution of measures of the state of network elements in different time periods; (iii) develop and test knowledge-based human-computer interfaces to help the users to specify the data marts to address specific analysis needs; (iv) exploit inference techniques to help in the specification of data marts and information analysis; and (v) test the proposed model for analyzing information of complex networks in other domains, like the traffic.

References


