

# Spatialization Methods: A Cartographic Research Agenda for Non-geographic Information Visualization

**André Skupin and Sara Irina Fabrikant**

**ABSTRACT:** Information visualization is an interdisciplinary research area in which cartographic efforts have mostly addressed the handling of geographic information. Some cartographers have recently become involved in attempts to extend geographic principles and cartographic techniques to the visualization of non-geographic information. This paper reports on current progress and future opportunities in this emerging research field commonly known as spatialization. The discussion is mainly devoted to the computational techniques that turn high-dimensional data into visualizations via processes of projection and transformation. It is argued that cartographically informed engagement of computationally intensive techniques can help to provide richer and less opaque information visualizations. The discussion of spatialization methods is linked to another priority area of cartographic involvement, the development of theory and principles for cognitively plausible spatialization. The paper distinguishes two equally important sets of challenges for cartographic success in spatialization research. One is the recognition that there are distinct advantages to applying a cartographic perspective in information visualization. This requires our community to more thoroughly understand the essence of cartographic activity and to explore the implications of its metaphoric transfer to non-geographic domains. Another challenge lies in cartographers becoming a more integral part of the information visualization community and actively engaging its constituent research fields.

**KEYWORDS:** Visualization, spatialization, cartography, dimensionality, self-organizing maps, multidimensional scaling, spatial cognition, human-computer interaction

## Introduction

A number of principal approaches have been put forward during recent years to give people the means for making sense of large, complex, and often unstructured data repositories. The problems encountered are shared across many knowledge domains. This has led to the development of distinct cross-disciplinary approaches, which draw on the accumulated knowledge of different academic traditions. For example, it would be hard to discuss current data-mining efforts without considering the role of traditional statistical inference. Likewise, one cannot ignore the influence of the vector-space model (Salton 1968) on modern knowledge discovery tools. It is surprising then that while mapping metaphors have long been popular in information visualization, decades of cartographic research—not to mention the broader cartographic tradition—have often been all but ignored (Card et al. 1999). Arguably, cartographers and geographers should be faulted more than anyone else for failing to engage the interdisciplinary information visual-

---

**André Skupin**, Department of Geography, University of New Orleans New Orleans, LA 70148. Tel: (504) 280-7157; Fax: (504) 280-1123. E-mail: <askupin@uno.edu>. **Sara Irina Fabrikant**, Department of Geography, University of California, Santa Barbara, CA 93106. Tel: (805) 893-5305. Fax: (805) 893-3146. E-mail: <sara@geog.ucsb.edu>.

---

ization community by demonstrating the relevance of their accumulated expertise. While computer scientists have been the most active contributors to information visualization research, and the institutional infrastructure is dominated by IEEE and ACM<sup>1</sup> activities, information visualization has remained an open, inclusive, interdisciplinary research activity.

Among cartographic research into non-geographic information visualization one can distinguish two strands of activities. Some cartographers are engaged in the interpretation and transformation of specific computational approaches in the light of cartographic tradition and informed by geographic information science (Skupin 2000, 2002a; Skupin and Battenfield 1996). In contrast to this *computational* perspective, the *cognitive* approach emphasizes the user side of

---

IEEE = Institute of Electrical and Electronics Engineers; ACM = Association for Computing Machinery.

spatialization. It aims at providing visualizations that function in accordance with what we know, or would like to know, about human perception and cognition of geographic space and its visual representations (Fabrikant 2001a; Fabrikant and Buttenfield 2001). The two perspectives are complementary, with geographic information science providing for a synthesis that matches geometric primitives against the cognitive categories that underlie our understanding and representation of space (Couclelis 1998; Fabrikant and Buttenfield 2001; Skupin 2002b).

Some of the influences behind cartographic spatialization are distinctly geographic; others are shared with a number of fields, such as cognitive science, linguistics, information science, and human-computer interaction (HCI). Within geographic information science, spatialization is most closely associated with the geographic visualization rubric (Buckley et al. 2000), but it also shares some common interests and methods with geographic data mining and knowledge discovery (Miller and Han 2001). Among specific geographic influences, the First Law of Geography (Tobler 1970) is particularly noteworthy. It boils down to the observation that everything is related to everything else, but closer things are more closely related than distant things. This principle played a role in the choice of multidimensional scaling for text visualization (Skupin and Buttenfield 1996), and it inspired ongoing efforts to uncover the cognitive underpinnings of spatialization (Fabrikant et al. 2002). Apart from visual depictions that one would now rightfully call spatializations (Goodchild and Janelle 1988; Li 1998) there are related efforts by geographers dealing with the mapping of cyberspace (Dodge and Kitchin 2001), investigation of specific methods commonly used for spatialization (Lloyd 2000), and use of spatialization as an alternative tool for analyzing human subject tests (Mark et al. 2001).

This paper gives an overview of a number of issues relevant to a successful engagement of information visualization by cartographers. It argues that cartographers can contribute to spatialization efforts at all levels, from making an informed choice among the various dimensionality reduction techniques to the development of cognitively plausible, interactive visualizations.

## Data for Spatialization

Information visualization is potentially applicable to a large variety of data. However, depending on the characteristics of a data set, very different approaches can or must be pursued in order to turn raw source data into a visual form. The

degree to which structural information is explicitly encoded is one of the dominant aspects influencing the choice of pre-visualization manipulations. There are considerable differences in how structured, unstructured, and semi-structured data may have to be processed before visualization techniques can be applied. This is significant enough to distinguish data types accordingly.

## Structured Data

Creators of traditional cartographic maps and modern geographic visualizations (as well as those engaged in the broader area known as scientific visualization) are familiar with the use of structured data that are stored in the form of database tables. These contain distinct observations for a given number of variables. No guesswork is necessary as to where one observation ends and another begins, or which values are associated with which variables. Structured data are accessible to spatialization routines in a fairly direct manner. Standard statistical preprocessing methods can also be applied, such as scatter plots or multidimensional scaling (MDS) layouts. Even geographic data (for example, multivariate census data) can be spatialized in order to explore characteristics of geographic objects in high-dimensional attribute space rather than geographic space.

## Unstructured Data

Many data that are of particular interest to information visualization have for the longest time only existed in an unstructured form, making their computational analysis a difficult proposition. Free-form text is a prime example of this. Most Web page content falls into this category. Hypertext can be analyzed and ultimately visualized according to its link structure (Girardin 1995), and similar analysis can be performed whenever explicit or quasi-explicit links between documents are encountered, for example in co-citation analysis (Chen and Paul 2001). However, *content* analysis depends on being able to further dissect elements *within* documents. Dealing with Web page content one faces the same hurdles as with most text content published in electronic form, such as conference proceedings. Much preprocessing is necessary before such data are suitable for computation of distance measures and the like. One needs to extract meaning-bearing elements from the text as well as assemble document metadata for further analysis. The necessary analytical approaches range from long-established information retrieval principles (Salton 1968) to methods for automatic

concept extraction (Chen et al. 1994). Much of the spatialization work done by cartographers has focused on using text data (Fabrikant 2000; Fabrikant and Buttenfield 2001; Skupin 2002a; Skupin and Buttenfield 1996, 1997).

## Semistructured Data

In recent years, an approach has emerged that provides mechanisms for making data self-descriptive in order to better support data exchange and automated analysis (Suciu 1998). Known as semistructured data, this approach employs schemas that act as models describing how data are structured. For example, in the semistructured storage of conference abstracts, one could divide individual abstracts into such components as the title, author information, abstract text, and keywords. The resulting data set would include information about its structure. In the spatialization context, this makes it easier to extract data from which high-dimensional distance models can be computed. The extensible markup language (XML) is by far the most prominent medium for implementing such data.

In the future, semistructured data will become the main form of source data when dealing with such diverse inputs as conference abstracts, news articles, or bibliographic entries. Legacy unstructured data will have to be converted into semistructured form, but capture of new data will be increasingly streamlined. For example, visualizing the geographic knowledge domain on the basis of conference abstracts is easier when participants of the annual AAG meeting submit papers through a form-based interface, as has recently been the case. Upon filling in the online submission form, its content can immediately be channeled into the appropriate structural elements, in accordance with a given schema.

## Methods for Dimensionality Reduction and Spatial Layout

Dimensionality reduction represents one of the most challenging tasks in any spatialization procedure. Cartographers are keenly aware of some of the issues that arise in this process, as reduction from a curved two-dimensional surface to a 2D plane is the basis of most cartographic depictions. They have developed guidelines for matching user requirements with projection types and devised numerous methods for communicating the inevitable distortions (Mulcahy and Clarke 2001). Deriving a suitable low-dimensional geometric configuration from a high-dimensional data set is likely to increase in difficulty with the

increase of dimensions involved in the transformation process. Database candidates for spatialization may contain just a handful of dimensions, for instance data containing the results of cognitive experiments (see Mark et al. 2001 for an example), or several hundred dimensions when dealing with large archives of indexed text documents. All map projections can in principle be reduced to a computation involving a function of latitude ( $\phi$ ) and longitude ( $\lambda$ ), some assumption made about the Earth's size ( $R$ ), and the desired scale ( $S$ ):

$$X = S * R * f(\phi, \lambda)$$

$$Y = S * R * f(\phi, \lambda)$$

The principles and algorithmic implementations for dimensionality reduction in spatialization are far more varied than this. In-depth understanding of one particular projection technique, e.g., multidimensional scaling, does not equip a researcher with the necessary understanding of the workings of another technique, e.g., self-organizing maps. Another major difference of spatialization compared to map-projection methods is that a feature's geographic dimensions (e.g. longitude, latitude and altitude, or width, length and depth) are physical properties established by a chosen frame of reference. These intrinsic dimensions have a meaningful order (e.g., altitude cannot be substituted by longitude), thus determining explicitly any feature's absolute position on the Earth's surface. In spatialization, however, dimensions are rarely intrinsically linked to objects, but are extrinsically assigned to establish relative locational relationships to other objects in the spatialization.

The task of choosing a spatialization technique is made harder by the fact that each method requires preparatory computations specific to the chosen technique in order to deal with its own set of peculiarities. Techniques can differ dramatically in terms of such desirable properties as scalability, incrementality, and robustness. For example, the self-organizing map (SOM) method is applicable to data sets containing very large numbers of observations and/or dimensions, while multidimensional scaling (MDS) is of little use for such data. All this makes it challenging to objectively compare different spatialization methodologies.

What follows is a discussion of some of the more popular methods employed in spatialization. While *dimensionality reduction* is an appropriate collective term for such methods as multidimensional scaling or self-organizing maps, some of the discussed methods should rather be referred to as *spatial layout* techniques. Such techniques are less concerned with

the preservation of high-dimensional proximities and more with an optimal use of available display space towards interactive use and reduced graphic complexity. Tree maps and some spring model implementations fall into this category.

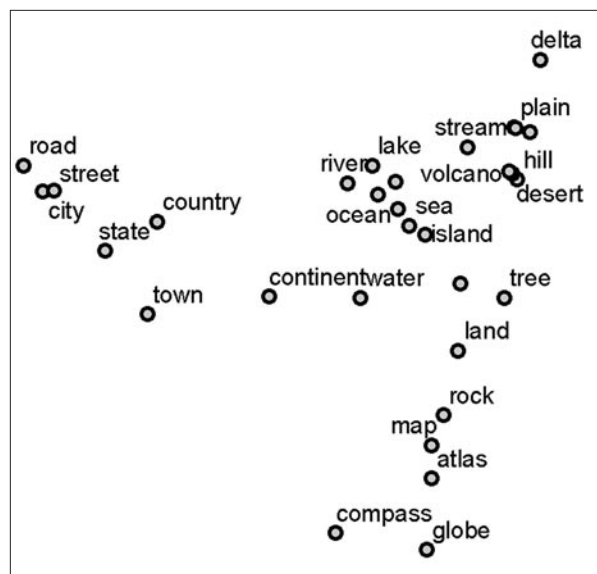
## Multidimensional Scaling

Multidimensional scaling (MDS) has historically been among the most popular methods for dimensionality reduction. A large number of variations have been proposed over the years, including metric and non-metric approaches (Kruskal and Wish 1978). The most commonly used MDS implementation is ALSCAL (Alternating Least Squares Scaling), which is the underlying method for MDS encountered in some statistical packages, such as SPSS.

Immediate input to MDS is always a dissimilarity matrix computed from a set of observations. Depending on the proximity measure chosen for the computation of this matrix, very different MDS solutions will result (Skupin and Buttenfield 1996). Throughout the preprocessing and scaling stages, input observations are treated as discrete objects existing in an  $n$ -dimensional vector space, where  $n$  represents the number of the perceived or existing attributes of all objects in this space. An MDS solution is a geometric configuration with low-dimensional coordinates for each individual object. In other words, an MDS point configuration is a spatial projection of the object's attributes, and when depicted, it can be called a map. Almost without exception, the derived point configurations are displayed with point symbols and associated labels (Figure 1).

It is rare to find solutions that process a configuration of MDS points further into interpolated surfaces or other derived forms. Rooted in psychometric research, and typically comprising input data of fewer than a couple dozen dimensions, MDS solutions have traditionally been derived in two dimensions for easy graphic depiction. However, technically, the method can be used to create solutions of higher dimensionality. For example, ALSCAL will allow a choice of one to six dimensions for the output configuration.

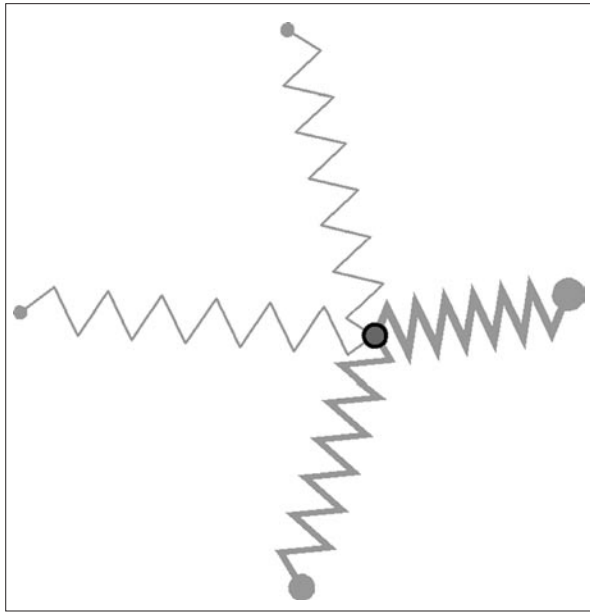
Multidimensional scaling was also an early favorite in spatializations done by cartographers and geographers. Tobler (1973) made early comments regarding the relationship between MDS and survey approaches to trilateration. Goodchild and Janelle (1988) used MDS in their extensive analysis of academic geography to map out research areas within the discipline. Skupin and Buttenfield (1996, 1997) used MDS to visualize news articles and noted the affinity of similarity-based mapping to the First Law



**Figure 1.** Typical visualization based on multidimensional scaling. Thirty-one terms elicited from human subjects during an investigation of geographic categories are visualized as a point configuration. The input data set contained five variables. (Mark et al. 2001).

of Geography (Tobler 1970). Cartographers have also used MDS to visualize the content of Web pages (Skupin 1998) and online catalog entries (Fabrikant and Buttenfield 2001). Outside of geography, the SPIRE (Spatial Paradigm for Information Retrieval and Exploration) project at the Pacific Northwest National Laboratory, and particularly its ThemeScapes product, received a great deal of attention (Wise et al. 1995). The core of its approach to dimensionality reduction is a variation on the MDS theme called the Anchored Least Stress (ALS) method, which aims at reducing the computational complexity of traditional MDS-based text visualizations (Wise 1999).

Multidimensional scaling remains a viable method for many data sets that have a limited number of objects and attributes. A number of issues could be further investigated using a cartographic approach. For example, virtually non-existent are visualizations of the distortions introduced by MDS. In other words, it is difficult to assess the degree to which high-dimensional proximities are preserved in a low-dimensional, geometric configuration. This is also the case for most of the other spatialization methods discussed in this paper. Another worthwhile cartographic approach relates to the visualization of locational object uncertainty due to input parameter modifications during the pre-processing stages and for alternative MDS computations. In dealing with uncertainty inherent in MDS and other methods, valuable inspiration may be derived from previous work in GIScience and geovisualization on uncer-



**Figure 2.** In a spring model, the location of individual observations is established through modeling of spring-like forces. Schematically shown here is the mapping of a single observation based on four variables, each with its own fixed origin.

tainty modeling (e.g., Zhang and Goodchild 2002) and uncertainty portrayal (e.g., Van der Wel et al. 1994; Clarke and Teague 2000).

## Spring Models

Underlying the spring model is a conceptualization of objects as nodes in a planar graph, with relationships between objects manifested as spring-like forces (Eades 1984; Kamada and Kawai 1989). As the method attempts to arrange nodes in a two- or three-dimensional space, the strength of springs gets adjusted through an iterative process, leading to an oscillation of nodes towards a minimum-energy configuration. In its simplest form, a spring model configuration can be constructed by mapping  $n$ -dimensional observations with respect to  $n$  fixed origins as shown in schematic form in Figure 2. However, the term spring model is collectively used for a heterogeneous group of techniques that are all based on force-directed placement.

Traditional graph drawing provided a major initial impetus to the development of spring models, and graphic drawing remains their major application. Graph drawing has traditionally been employed to depict microchip connections, software engineering diagrams, or any type of computing networks. Its goal is to optimize drawing criteria in order to develop an aesthetically pleasing, well structured graph with nodes and links in two dimensions. Much attention is paid to prevent link crossings and to

ensure optimal use of available display space, evenness of edge lengths, and other quantifiable aesthetic layout criteria.

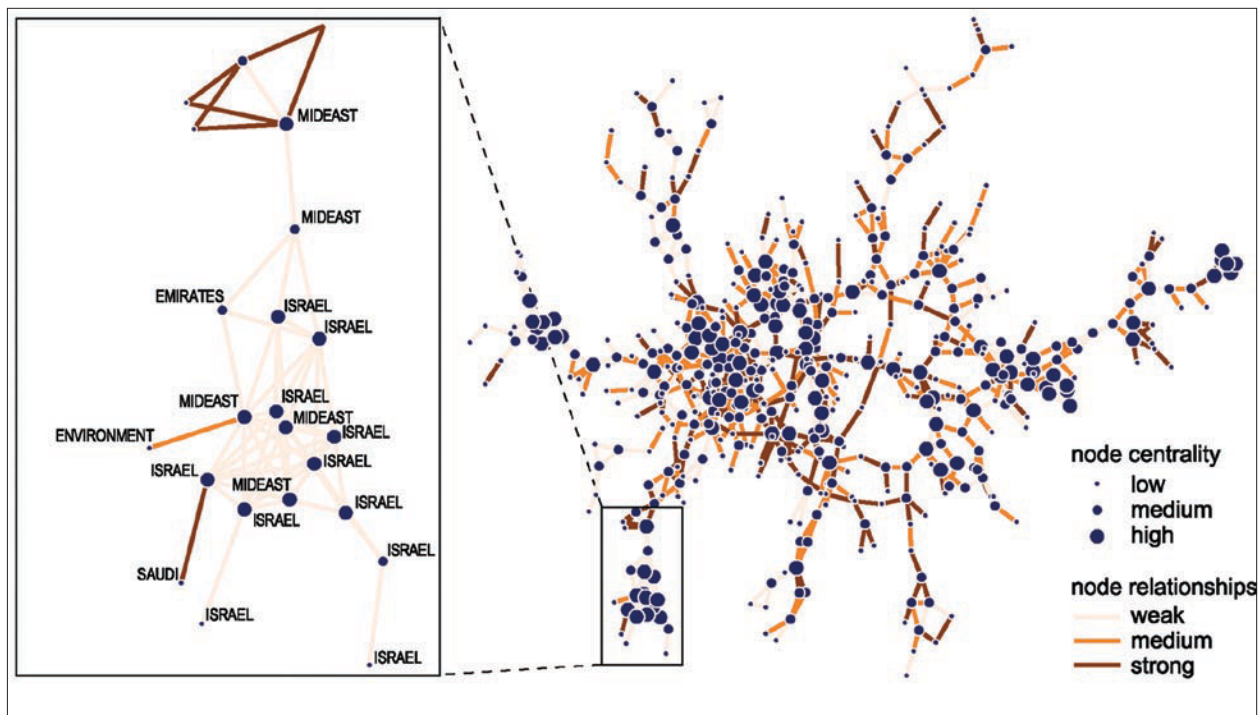
Alternatively, spring models can be applied to proximity data, such as the semantic similarity of journal papers. In this case, spring forces are not derived from a pre-determined object topology but emerge from the analysis of pair-wise object similarities. In this form, spring models are a valuable alternative to MDS, particularly when dealing with very large and high-dimensional databases (Chalmers 1996), because spring models are computationally more efficient than multidimensional scaling. The resulting two-dimensional geometric configurations consist of individual point locations, which can be visualized accordingly.

Spring models, similar to MDS, can be combined with other types of data reduction methods (e.g., clustering, network scaling) to generate richer spatial representations. For example, Fabrikant (2001b) used a spring model to create a two-dimensional point configuration of Reuters news stories and then combined it with a pathfinder network scaling solution to depict a semantic flow map with news stories represented as nodes in the network (see Figure 3).

## Pathfinder Network Scaling

Pathfinder scaling, also known as pathfinder network scaling (PFN), is a popular method for depicting edges or links between nodes of a graph representation (Schvaneveldt 1990). Starting from a proximity measure between nodes, PFN solutions aim to uncover the most essential links among the nodes in a network (Chen and Paul 2001). The method is essentially based on a minimum spanning tree algorithm that aims at deriving, from input proximities, a minimum spanning tree (of many possible ones), preserving the most salient links between the nodes. The resulting visualizations combine the complete set of nodes with a limited number of links, aiming to guide the human eye to more easily recognize important node relationships (Figure 3).

The depiction of semistructured data with node-link representations is distinctly different from many MDS or spring model approaches in which point locations alone provide visual cues of the inherent data relationships. The PFN approach has been especially popular among information scientists working on knowledge domain visualizations, for example with author co-citation analysis (Chen and Paul 2001). As mentioned earlier, the network topology created by PFNs can be projected onto existing two-dimensional solutions derived from other kinds



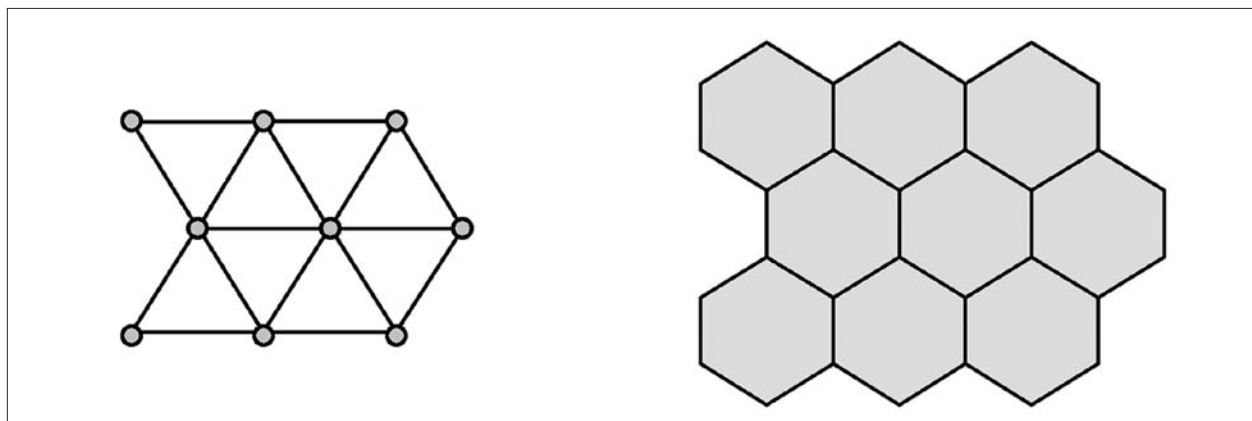
**Figure 3.** Pathfinder network scaling provides explicit representation of dominant semantic proximity relationships between individual documents in a high-dimensional Reuters news archive (redrawn from Fabrikant 2001b).

of point scaling methods. Recent empirical findings suggest that the provision of explicit visual links between point locations alters the perception and cognition of semantic proximities in 2D information spaces (Fabrikant et al. 2002), but one should use pathfinder scaling to this end only *if* the source data and the employed distance measures warrant such strong visual messages.

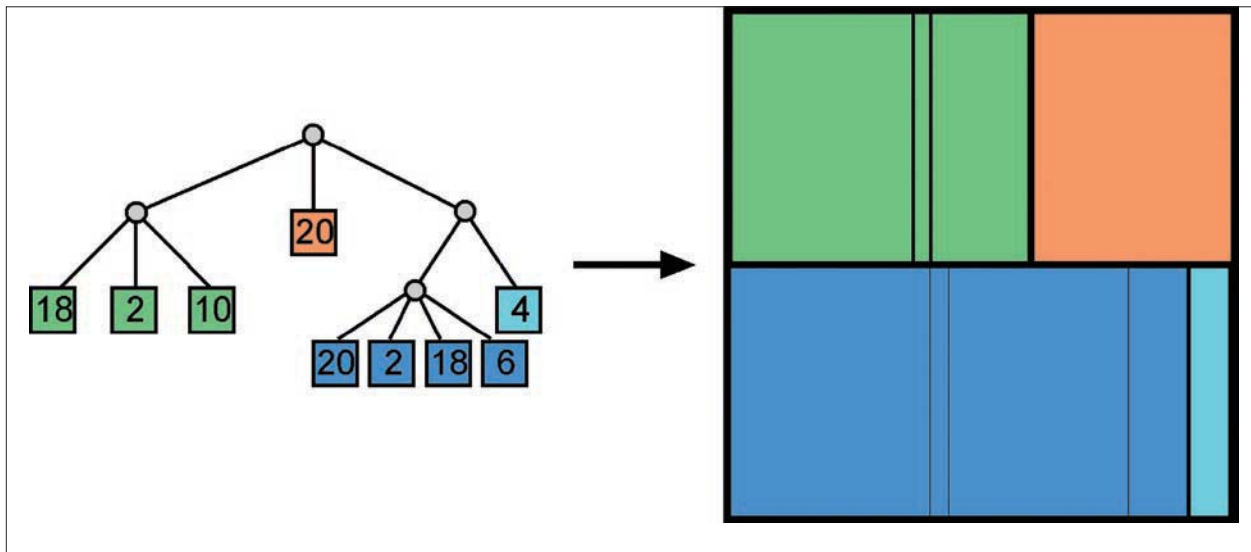
### Self-Organizing Maps

The self-organizing map (SOM) method (Kohonen 1995) has received much attention in recent years,

with applications as diverse as medical imaging, voice recognition, stock market analysis, and even artistic installations in museums (Legrady and Honkela 2002). Some uses of SOM now popular in cartographic spatialization, such as text document visualization, were first demonstrated a decade ago (Lin 1992). In general, geographic SOM applications tend to concentrate on data classification rather than information visualization tasks. This is somewhat surprising, given that its predominantly two-dimensional graphic form (as shown in Figure 4) lends itself to easy integration and visualization



**Figure 4.** A self-organizing map (SOM) is typically laid out as a two-dimensional artificial neural network. The regular spacing of neurons lends itself to raster visualization.



**Figure 5.** The treemap method takes a tree structure as input to tessellate a given display area. Area sizes are frequently varied in accordance with numeric attributes.

with GIS, taking advantage of available spatial data structures and spatial analytical capability.

For example, interactive side-by-side display of census data in geographic space and SOM attribute space has been demonstrated (Li 1998). Integration of large SOMs (i.e., at least several hundred neurons) into GIS is possible (Skupin 2000, 2002a) but tends to rely on very loose coupling of GIS and SOM components, which hinders efficient implementation. Closer integration of geographic software and SOM components has long remained elusive, but this may become easier with the availability of such visualization environments as GeoVISTA Studio (Gahegan et al. 2002), which provides JavaBeans for SOM training and display. Compared to other methods, SOMs scale up far better for very large and/or very high-dimensional data sets. While training times of up to six weeks on a six-processor system may, at first glance, seem excessive (and certainly preclude interactive use), one would be hard pressed to find other methods able to deal with data sets of several million documents (Kohonen et al. 1999).

The SOM method is likely to remain an important projection method in the future, providing cartographers with interesting representational challenges. A number of SOM visualization techniques have been proposed. One can distinguish between two major visualization categories. First, the trained SOM itself can be shown. Examples for this are visualizations of component layers, high-dimensional neuron clusters, and visual depictions of the distortions caused by SOM training. Commercial off-the-shelf GIS software is suitable for visualizing a trained, two-dimensional SOM. The regular lattice of neurons suggests the use of a raster data structure (Skupin 2002b). However,

hexagonal neighborhoods, in which each neuron is connected to six neighbors, have traditionally been used more frequently than square neighborhoods. Lacking support for hexagonal pixels means that individual neurons are best represented as polygons in GIS, as shown in Figure 4.

The second principal approach, and one that is more akin to typical neural network applications, uses a trained SOM as the basis for visualizing a set of observations that may or may not have been part of the original training data set. This typically takes the form of point visualizations, since individual input observations become associated with individual neurons from which coordinate pairs can be derived.

The relevance of GIScience expertise is not restricted to visualization issues. There are several interesting research questions to pursue in the area of spatial data models, generalization, and error assessment.

## Tree Maps

Among the techniques discussed here, the tree map method (Johnson and Shneiderman 1991) is arguably the one that is most widely known and popular outside the scientific community. Variations of the tree map method appear, typically under colorful names, in various commercial information visualization applications on the Internet. Popular examples are the mapping of Web space at “antarcti.ca” or the daily updated stock market map at “money.com.”

The success of the tree map method can largely be explained by its ability to express hierarchically organized data through map-like visual structures,



**Figure 6.** Illustrating the effects of neural network training through cartographic means. Locations outside of the landmasses were deliberately excluded from training, which leads to a contraction of the corresponding space portions. Areas near the edge of this SOM are also significantly distorted.

leading to depictions that combine the visual attractiveness of maps with the cognitively useful organization provided by a nested hierarchy. Another important factor for the success of tree maps is that they are technically less difficult to implement than the other techniques discussed in this paper. Human subject evaluations of tree maps have not yet provided sufficient empirical evidence on how people understand the invoked metaphorical distance-similarity mapping. As with other methods, insights into the usability of spatialized displays from a human-computer interaction perspective are only slowly emerging (Chen and Czerwinski 2000).

Tree maps are generally not used to visually uncover high-dimensional structures and relationships. Instead, existing tree structures are the input to tree-map algorithms, depicted in 2D. The tree map method is thus only applicable when a hierarchical structure has already been established. Examples are the directory structure of computer operating systems or the Web hierarchy established through the Open Directory Project (ODP). Tree maps tessellate a given display space, such that individual tree nodes occupy a certain proportion of the space based on their position and/or importance in the data hierarchy. Tree maps behave similar to area cartograms in that polygon size is often scaled to the magnitude of quantitative node attributes (Figure 5).

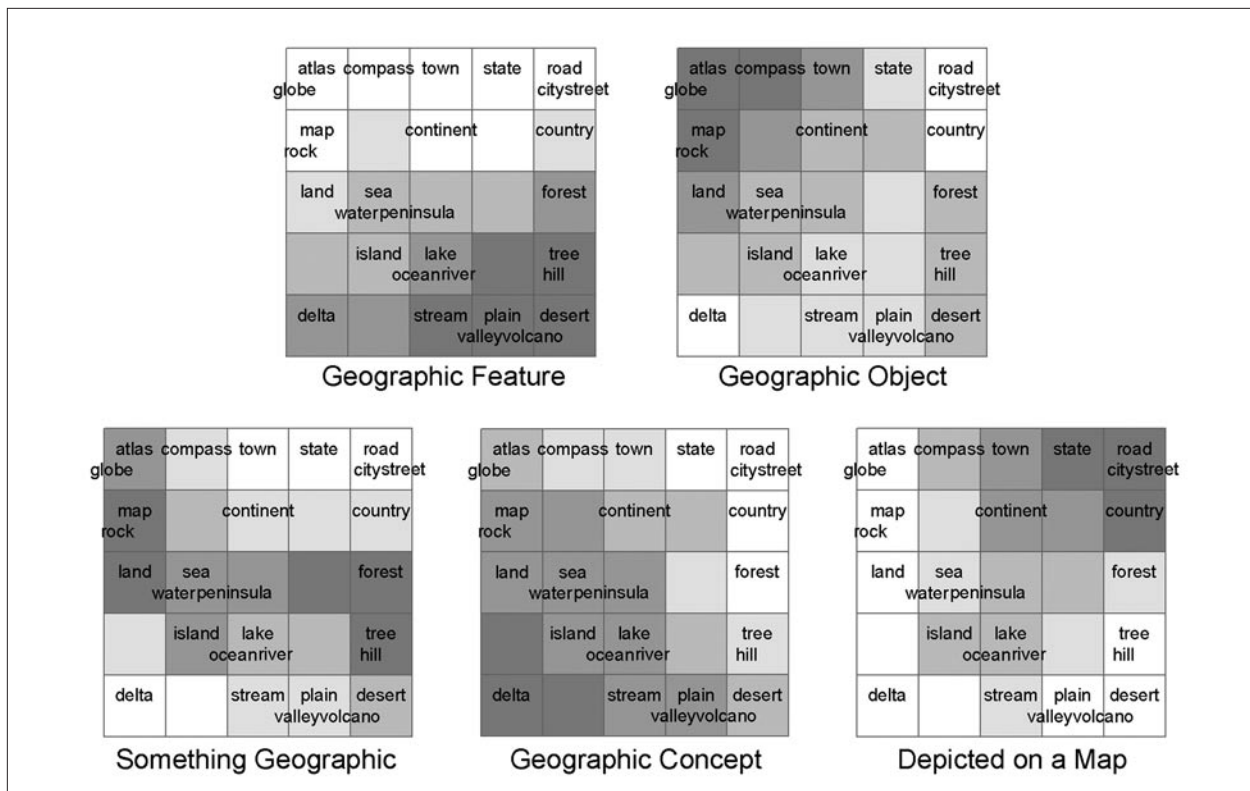
In an interactive setting, “scale change” is initiated by clicking on a particular polygon representing an information node in the tree. The tree portion below the selected node is visualized, and so forth. While

this sounds like a reasonable form of interaction, different tree-map implementations vary significantly in how scale changes are computationally implemented and visually conveyed. In particular, the relationship between different zoom levels is often implemented in surprisingly un-cartographic ways, as seen with the WebMap system ([www.webmap.com](http://www.webmap.com)). In that implementation, the tree map consists of the usual tessellation into a number of polygons, plus a terrain-like interpolation visualized through hypsometric tinting. Clicking on a polygon leads to an expanded view, which again consists of polygons and terrain interpolation. However, each of the nodes in the tree is visualized in isolation from all the other nodes. Each terrain view is based on its own isolated interpolation, which leads to the curious effect that clicking on neighboring polygons will result in terrain views that do not transition fluidly at their imaginary common boundary.

### Other Techniques

The list of techniques used for dimensionality reduction and spatial layout is too extensive to be comprehensively discussed in this paper. For example, one could easily include such traditional methods as principal components analysis (PCA), principal coordinate analysis (PCoA), correspondence analysis (CA), projection pursuit, or such newer developments as Isomap (Tenenbaum et al. 2000). Most of these techniques (as well as those discussed earlier) have evolved from dis-





**Figure 7.** Results of the human subject test visualized with a 5-by-5 neuron SOM. Five component planes or layers of the SOM are shown. Each component plane corresponds to one input variable. Labels indicate the terms associated with each neuron.

tinct academic traditions. For example, MDS has an unmistakable psychometric lineage, while the closely related technique known as Sammon mapping (Sammon 1969) is more associated with an engineering tradition. As a result, terminological and methodological distinctions between Sammon mapping and MDS are not always clear. Such difficulty of comparing and classifying different methods is common when one is working in an evolving, interdisciplinary field such as information visualization.

### Through the Eyes of GIScience

A GIScience-based approach can add a new dimension to the discussion of different scaling techniques. One could distinguish projection methods according to their underlying conceptualizations (e.g., discrete objects vs. continuous fields) and employed geometric primitives (e.g., points vs. polygons) (Skupin 2002b). This helps to explain why certain scaling techniques tend to be associated with particular visualization forms. For example, MDS employs a conceptualization of high-dimensional space as being mostly empty, except for the existence of a finite number of

discrete information objects. This object conceptualization is felt throughout the stages of typical MDS analyses. In contrast, the SOM method conceptualizes input observations as samples from a high-dimensional, continuous, information field. The implications of this conceptualization propagate through SOM-based information processing and are manifested as raster-like visualizations to the user.

Regardless of the conceptualization approach, any projection from a high-dimensional information space to a low-dimensional representational space always leads to distortions, either through space contraction or expansion. Knowledge of the distortion characteristics of the various dimensionality reduction methods ought to influence the choice of a particular technique for a specific information need. However, not only are comparative analyses of the distortion characteristics of different methods rare, but graphic representation of distortions in spatialization is also remarkably absent. Cartographers and GIScientists can take on a leading role here, by drawing on cartographic approaches to the investigation of map distortions, inspired by Tissot indicatrices, displacement vectors, and other existing methods (Tobler and Wineberg 1971; Tobler 1976; Mulcahy and Clarke 2001).

There is also place for critical reflection on existing distortion depiction methods, such as the popular U-matrix visualization method for SOMs (Ultsch 1993). While it is billed as a cluster visualization technique, it actually visualizes contraction and expansion effects of SOM training, i.e., the distortion resulting from fitting object relationships in high-dimensional space into only two dimensions using a topology-preserving approach.

Cartographic approaches can also be applied to graphically convey effects in the solution space, when choosing particular techniques or when modifying pre-processing parameters and input settings. For example, with the SOM method, one could visually demonstrate the importance of including representative samples in the training of an artificial neural network. Figure 6 shows a world map based on a SOM whose training data set only contained points inside the landmasses. Topological relationships of connected countries are preserved, but regions without training data (i.e., the oceans) undergo contraction. As a result, the Atlantic Ocean becomes a mere stream separating the Americas from Europe and Africa.

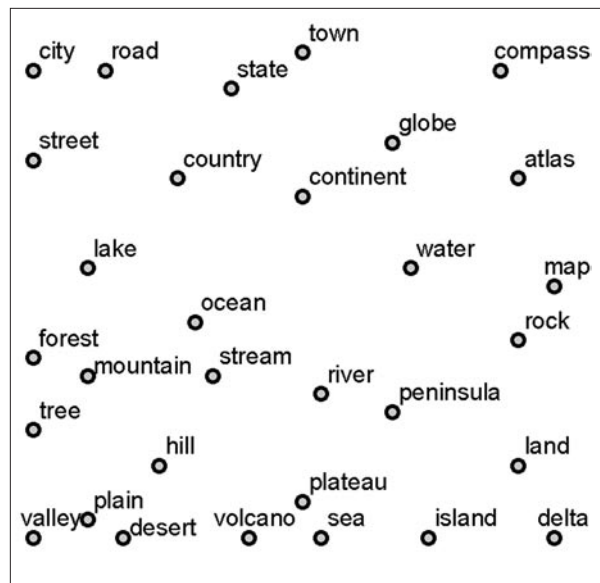
Resolution dependency, another peculiarity associated with the SOM method, can also be illustrated with cartographic means. In Figure 6, SOM granularity is such that land areas immediately bordering the Straits of Gibraltar become associated with the same neuron. This prevents the Mediterranean Sea (shaded in gray like the landmasses) from being connected to the Atlantic Ocean.

## On the Role of Transformation

Unless otherwise noted, all the visual examples in this section were derived from a human subject test dealing with geographic ontology. Refer to Mark et al. (2001) for a detailed description of the experiment and its implications. The empirical results consist of 31 terms that were elicited with respect to five differently formed test questions.

Design decisions during information visualization tend to be intimately linked to the characteristics of the projection technique. Thus, discrete objects entering an MDS procedure will usually lead to point visualizations (see Figure 1), while the field conceptualization of SOM leads to raster-type visualizations (Figure 7).

There are a number of reasons why one may want to further transform the prototypical displays associated with certain projection techniques. It may be necessary to perform geometric transformations in order to enable certain visualization methods that may be required by a specific information need. For example, terrain-type visualization might require interpolation

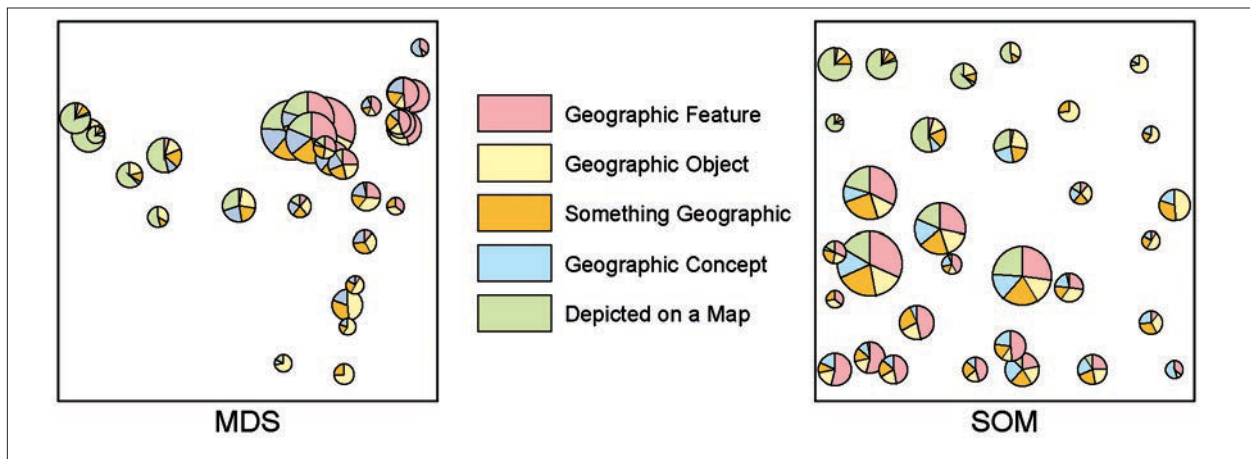


**Figure 8.** Results of the human subject test visualized on the basis of a 30-by-30 neuron SOM. Due to the large number of neurons, each captures not more than one term, allowing the creation of a unique point location for each of the 31 terms. This makes comparison with the MDS configuration easier.

based on some carefully chosen attribute. Assuming that one is dealing with two-dimensional input geometry, the full complement of data transformations in the GIScience arsenal can become relevant here. Once the desired geometric model is established, the considerations and methods of traditional cartographic design and contemporary geographic visualization are also directly applicable.

## Visualization Method

The most obvious need for transforming a geometric configuration produced by a chosen dimensionality reduction and spatial layout technique derives from the data model requirements of the visualization environment used for data exploration. For example, if discrete point symbols are required for a particular information need, the question is how can one derive a point configuration from a trained SOM that would typically be visualized as a surface? In short, one has to find the neuron that matches a given object closest in high-dimensional attribute space and use that neuron's low-dimensional location for depiction. The specifics of this transformation procedure are dependent on the desired SOM resolution. With a high-resolution SOM, neuron locations may be directly usable. Figure 8 shows a point visualization derived from a 30x30 neuron lattice comprising 900 neurons; this



**Figure 9.** Two visualizations based on the MDS and SOM configurations shown previously. Pie charts are constructed from the same five variables used to create both configurations. Visual similarity of charts corresponds to geometric proximity. Note how the SOM method makes full use of the available display space, at the danger of distorting relative distances.

is more than enough to determine a unique location for each of the 31 input objects.

With a different object-to-neuron ratio, further transformations may be necessary, since a single neuron might become associated with multiple objects. Skupin (2002a) describes the mapping of 2220 objects using a 60x80 neuron lattice (4800 neurons). Unique point locations are determined for each spatialized object and placed randomly within the polygon associated with the closest matching neurons.

These and other space transformations inspired by GIScience are also useful for supporting meaningful visual comparison between different projection methods. For example, they can help to compare MDS solutions (Figure 1) to those created by the SOM method (Figure 8). Once two spatializations are in a common representational framework, including geometric form and symbolization, the results of each procedure can be visually inspected and compared (Figure 9).

Some objects in the MDS solution are part of dense clusters, making it hard to distinguish them in both the labeled point map (Figure 1) and the pie chart map (Figure 9). Point features in the SOM-based solution are spread out more evenly across the display area (Figures 8 and 9). These differences can be explained by the fact that MDS attempts to preserve input proximities in output distance relationships, while SOM is more focused on retaining topological relationships. As a result, empty attribute space portions in the SOM are contracted, potentially causing quite dissimilar objects to be positioned in close proximity, as seen near the left edge of Figure 8 (e.g., “street” and “lake”).

Preliminary empirical results (Fabrikant et al. 2002) indicate that explicit linear connections (links) or linear separations (boundaries) help modify potentially problematic human perception and cognition issues of straight-line (metric) distance relationships as implied by the First Law of Geography. Given the inevitable distortions encountered when projecting high-dimensional data into low-dimensional representations, it then appears important to mitigate distortion artifacts that may lead a user to false conclusions, by providing additional visual cues.

For example, as shown in Figures 10 and 11, one could combine a complete hierarchical clustering solution with a low-dimensional point configuration. Point objects in Figure 11 are linked in the two-dimensional display space according to their position in a hierarchical clustering tree (Figure 10). Line thickness corresponds to different proximity levels at which objects merge to form clusters in high-dimensional space. Thicker lines indicate tighter high-dimensional clusters that are merged at lower levels in the clustering tree.

Space transformation approaches and subsequent visualization of different spatialization solutions follow the call by the geovisualization community to generate many graphic realizations for a single data set to support abductive knowledge discovery, rather than concentrating on communicating a message with “the single optimal 2D map” (MacEachren and Kraak 2001). For example, three-dimensional landscape visualization methods based on raster interpolation are possible transformation techniques when an additional (understood) attribute dimension is added to an existing 2D point configuration. The choice of a particular interpolation method may be

dependent on a set of diverging goals, such as fitting an appropriate visualization to specific data characteristics (e.g., discrete vs. continuous data objects) to a chosen projection method (discrete vs. continuous space conceptualization), to user's cognitive abilities, or to specific information needs.

Figure 12 shows four different 3D landscape visualizations derived from an initial 2D spring model that treats input data as discrete observations and outputs coordinate pairs for each of them. Density surfaces are derived from the two-dimensional point configuration by means of spatial interpolation, where new data are created to fill the void between discrete observations. Compared to the stepped density surface (upper right) based on Voronoi polygons derived from the same discrete point locations, is the more natural looking continuous density surface (upper left) more adequate for information exploration at lower levels of detail? Does the stepped surface more appropriately convey the discreteness of the input data and the abrupt thematic change between cluster boundaries? Does it make sense to combine the continuousness of a natural landscape (thus emphasizing the experiential effect of the spatial metaphor) while at the same time preserving the discrete nature of the input data by means of pycnophylactic interpolation (Tobler 1979)? As the lower left panel illustrates in Figure 12, regardless of the motivation for the use of the pycnophylactic transformation technique, this method has a smoothing effect, which may be utilized to depict data at different levels of detail.

Even if conflicting depiction goals may not be resolved with one spatialization, current geovisualization approaches and tool developments provide the necessary visual exploration environments to dynamically inspect the range of possible alternatives and empower the user to proactively inspect the properties of diverse methods for dimensionality reduction and spatial layout.

## Generalization

Graphic and semantic complexities are prime foci of information visualization research that lend themselves to in-depth involvement of the cartographic community. Cartographers have had to develop numerous generalization methods to accommodate a wide range of map scales. While cartographic generalization is far from being completely understood, not to mention automated,

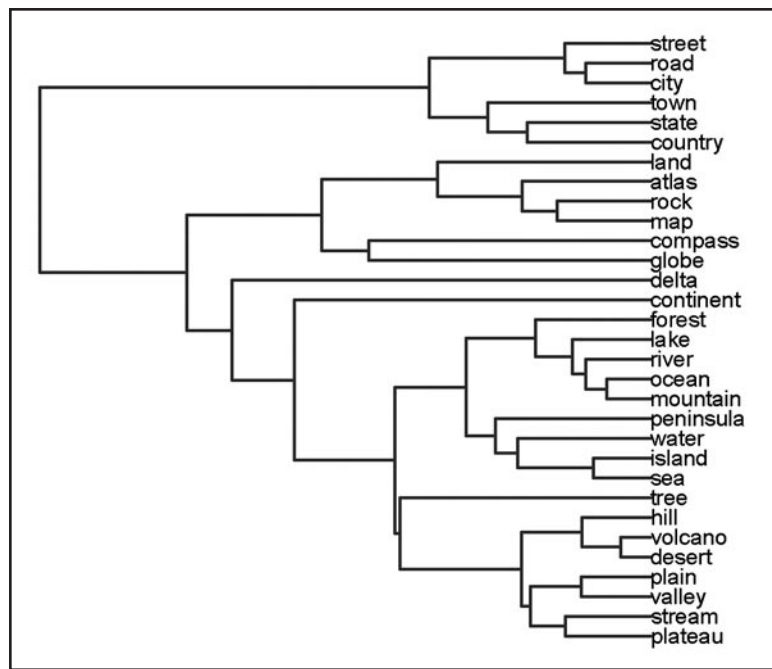
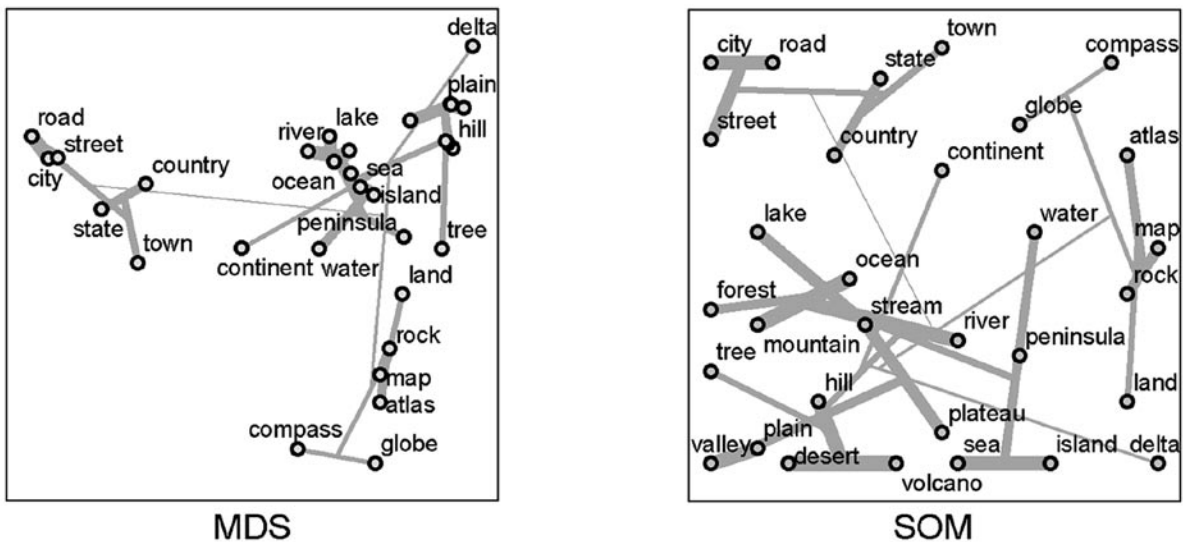


Figure 10. Dendrogram of a hierarchical clustering solution for 31 terms.

one would be hard pressed to find a community more devoted to this kind of issue, or one that has amassed more empirical knowledge. The cartographic approach to generalization also continues to distinguish itself from other purely graphic complexity approaches, such as the popular level-of-detail (LOD) philosophy in computer graphics, by attempting to address both geometric and semantic aspects of map complexity.

Recent studies demonstrate the relevance of the *scale* notion for non-geographic information visualization (Fabrikant 2001a; Fabrikant and Buttenfield 2001). This strengthens the argument for the incorporation of traditional cartographic generalization into scalable information visualization environments. Hierarchies are particularly useful vehicles for implementing cartographically informed visualization. Computed hierarchies, such as those derived through hierarchical clustering, have been proposed for cartographic generalization of geographic data (Ormsby and Mackaness 1999). Early cartographic attempts at scalable information visualization through integration of hierarchical clustering with MDS suffered from the limited applicability of MDS for large, high-dimensional data sets (Skupin 1998). Recent experiments in combining hierarchical clustering with SOM (Skupin 2002a) or spring-based methods (Fabrikant 2001b) appear more promising. Figures 13 and 14 show how a zoom operation might be derived by linking spatialization geometry to a cluster tree. Refer to Skupin (2002a) for a detailed description of how this multi-scale representation is



**Figure 11.** The hierarchical clustering tree shown in the previous figure is here projected onto the two point configurations created by MDS and SOM. Line thickness indicates the level in the clustering tree at which a merge occurred. This explicit indication of high-dimensional similarities may help to counteract the distortions introduced to varying degrees by either of the two methods.

derived, including the computation of scale-dependent cluster labels.

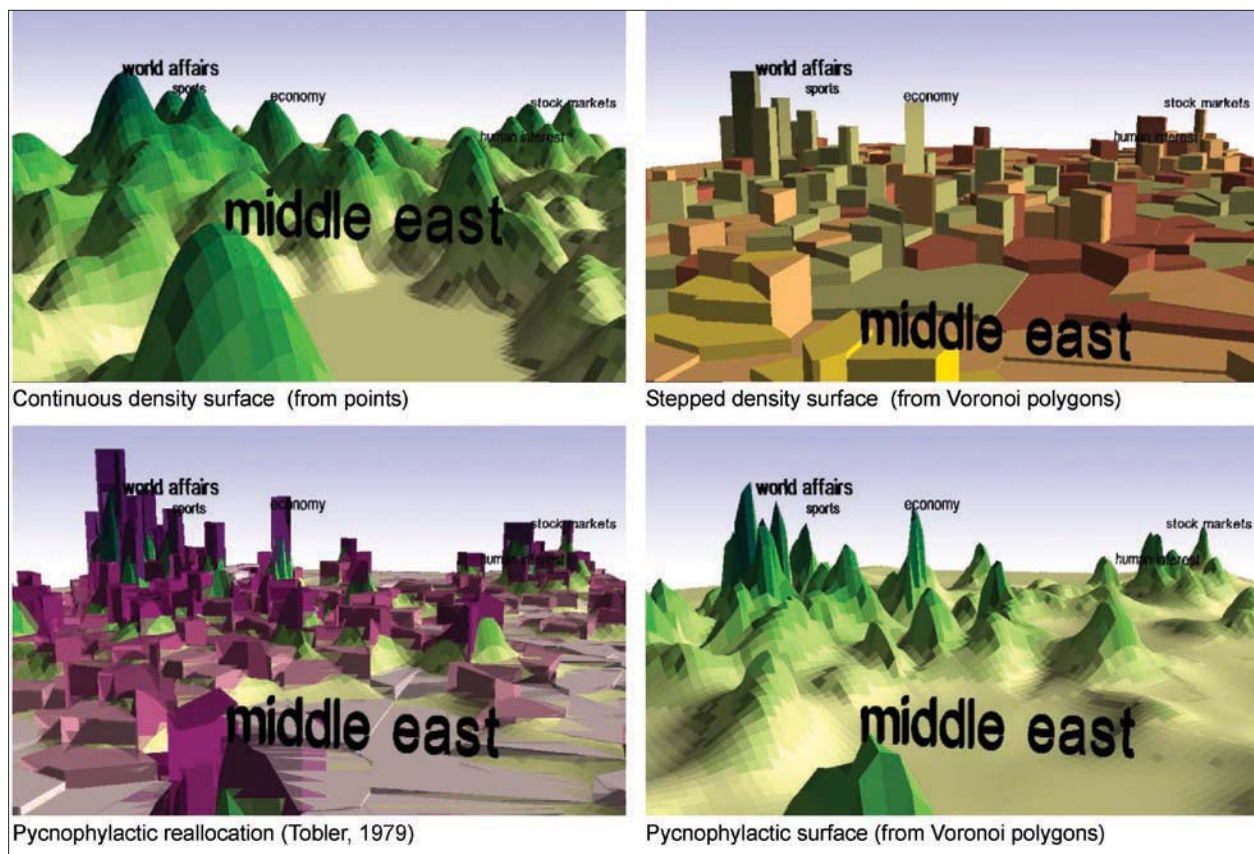
Ongoing cartographic research tasks in this area deal with computational and cognitive issues of complexity. On the computational side, issues of graphic density and object selection (Töpfer and Pillewizer 1966) are of interest. On the user side, much remains to be known about usability and usefulness of multi-scale visualizations, such as those shown in Figure 14.

Issues of graphic complexity have plagued information visualization researchers for many years. While some of the approaches put forward by the InfoVis research community amount to reinventions of the cartographic wheel, others are quite different from traditional cartographic and GIS solutions and should be critically engaged by the GIScience community. This refers particularly to the various distortion-based techniques (Leung and Apperley 1994), such as hyperbolic trees (Lamping and Rao 1996), and dynamic variations on fisheye views (Sarkar and Brown 1994) that cartographers previously explored in static form (Tobler 1973). Cartographic reflection on space-scale diagrams and related zoomable user interfaces is also overdue (Bederson et al. 1996; Furnas and Bederson 1995). Apart from the need for cartography's input to the further improvement of scale-dependent spatializations, most of the methods named above have also yet to be investigated by cartographers in terms of their suitability for geographic visualization.

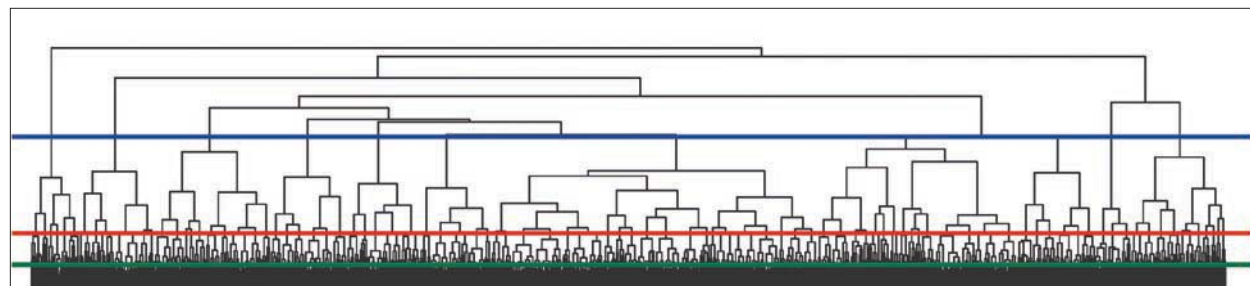
## Cognitive Considerations in the Design of Interactive Spatializations

Research opportunities for GIScientists and cartographers in spatialization are not restricted to computational techniques that produce meaningful spatialized geometries, visualizations, and methods of analysis. Possible research topics also encompass how information seekers may be able to more efficiently search visually and extract information dynamically from interactive spatialized displays, and thus make better sense of knowledge buried in large digital data archives. Spatialization visually summarizes and describes large data repositories and also provides opportunities for visual query and sense-making of large data collections. Improving knowledge discovery in data-rich environments is also a key concern in the GIScience community. For example, Geospatial Data Mining and Knowledge Discovery and Geographic Visualization have been identified as emerging research themes by the University Consortium of Geographic Information Science (Buckley et al. 2000; Buttenfield et al. 2000).

Recent research dealing with disseminating and accessing very large geographic data collections, including aerial photographs, satellite imagery, and digital and analog maps, also documents an increased use of content-based or semantic retrieval strategies (Castelli et al. 1998; Ma and Manjunath



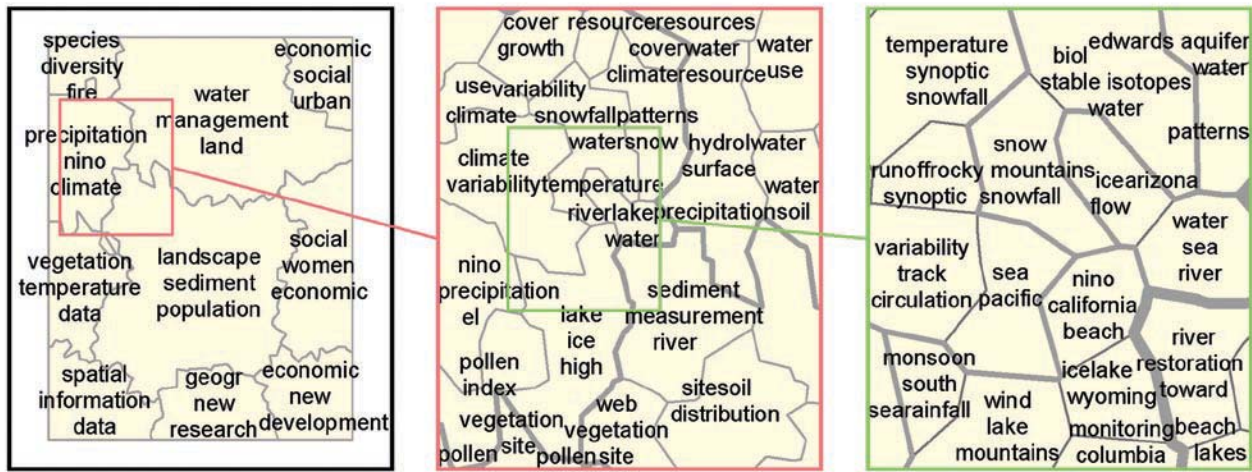
**Figure 12.** Density surfaces derived from a single, two-dimensional, spring configuration.



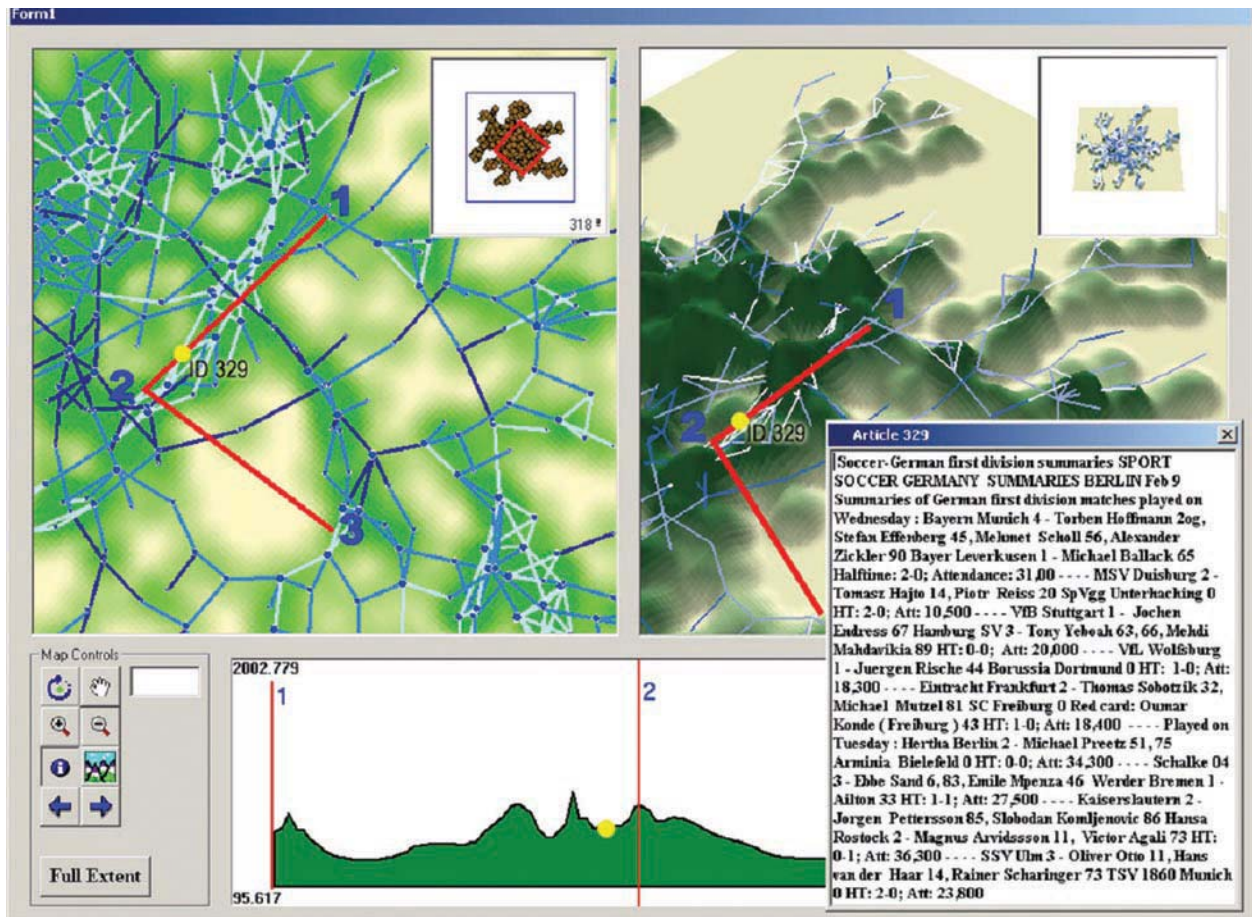
**Figure 13.** Hierarchical clustering tree with indication of three distance levels. Several thousand text documents were used to train a 60-by-80 neuron SOM consisting of several hundred component planes. A hierarchical clustering solution was then computed for the 4800 neurons in order to support scale-dependent visualization (Skupin 2002a).

1996; Manjunath and Ma 1996; Sheikholeslami et al. 1999). While retrieval systems differ in their level of abstraction (Chang et al. 1997), image-based query results are typically provided in graphic thumbnail form. This strategy might be effective with queries that return a manageable number of query results. However, it has been shown that users prefer to navigate within a clearly defined hierarchical semantic space (Chang et al. 1997). Transforming visual data archive content into semantic ontologies of visual information (Chang et al. 1997) will increase in importance as data archives are expected to grow exponentially.

Spatialization is based on envisioning spatial properties and requires the user's understanding thereof. Hence, it is important to study and apply cognitive principles for real spaces, which involve spatial reasoning, and communication about features, their spatio-temporal and thematic attributes, as well as the relationships among these objects in the real world. Preservation of geographic primitives and spatial principles in a spatialization allows interpretations about the content of the information space and places the transformation in a sound semantic framework. Figure 15 depicts an experimental spatialized user interface that is currently being developed and used



**Figure 14.** Three zoom levels in a visualization of conference abstracts. The merging of individual documents into regions is based on the hierarchical clustering tree shown in the previous figure (Skupin 2002a).



**Figure 15.** Spatialized user interface for exploring a news wire archive. The map view (upper left window) is linked to the 3D landscape view (upper right window), providing an information seeker with multiple perspectives of the same data set. A third window (bottom center) displays a *semantic* profile along a *semantic* transect line drawn in the 2D map and 3D landscape windows. The transect line is a spatialized query metaphor for identifying implicit cross-references between a source document (1) and several target documents (2 and 3). A user interested in European football results (e.g., German Bundesliga) has selected a document (ID 329) by mouse click. This news item is located in semantic proximity to a previously identified 'landmark document' (2). The content of the selected document (i.e., football match scores) is displayed in the document window (lower right).

for usability experiments, based on the geographic primitives' location, distance and scale.

Relative and absolute location provides a sense of a document's existence in the collection and of its semantic relationship with other documents in the archive. In Figure 15, semantic relationships between news stories are spatialized with the combination of a spring model and pathfinder network scaling. Each point in the display is linked to its original news story, which can be viewed and read while browsing the spatialized archive. A density surface was interpolated from the point configuration derived by the spring model. The surface model gives the information seeker a sense of the semantic density in the data space. A user can navigate in synchronized displays, e.g., in 2D (pan and rotate), or in 3D, with fly-throughs or walks. Coupling the location metaphor with distance, several documents may be cross-referenced ("brushed") by a linear connecting transect, or along the computed semantic network in the two-dimensional case, and simultaneously viewed in the spatialized 3D view. When navigating in a 3D space, a transect line might be represented by the shortest path between two points along the line of sight, or along a *semantic travel* path. Items falling along this path may be characterized as being more similar to one item (e.g., a specific *landmark* document used as the start of navigation) or another item (e.g., the destination document) (see Figure 15 for an example). Linked views at various scales are provided with inset displays. Small-scale inset maps provide a frame of reference (absolute location) and act as orientation aids for information seekers, when zooming deep into the data space (as shown in the larger windows in Figure 15). Small-scale views give the user a sense of the extent and size of the information space, even when they wish to zoom into higher levels of detail or navigate along paths at the bottom of deep *semantic* valleys where overviews of the landscape may be obstructed.

Content-based visual query research increasingly applies hierarchical ordering and clustering techniques based on functional distance or metric measures of semantic relatedness. Other approaches are intended to model perceptual similarity (e.g., similarity in shape, color, texture) between items in a large data archive (Healy and Jain 1996). In spatialization, documents within a given distance of a point of interest may form graphic clusters of related information. Mountains and valleys of documents structure the information space and allow exploration in linked windows in 2D and 3D. Information clusters may be nested hierarchically. Clusters can be explored at different levels of detail, introducing the concept of scale (Figure 14).

A recent study provides empirical evidence supporting the usability of spatialized views. The research included the creation and evaluation of a spatialization prototype to access a large document collection similar to the one depicted in Figure 15 (Fabrikant 2000; Fabrikant and Battenfield 1997). The design and implementation of spatialized interface components were based on three spatial concepts: distance (similarity), arrangement (dispersion and concentration), and scale change (changing level of detail). Empirical evidence was collected on the effect of people's background and training on metaphor association, as well as the effect of representational variables such as data type, dimensionality, color and shape. The study showed that people associate (1) interpoint distance with the concept of document similarity in a document collection; (2) graphic clusters representing the information content and structure of a digital collection with concentration of related documents; and (3) graphical change in resolution (zoom-in) with different levels of detail in a document collection (hierarchical order). One of the most striking results in this study is that metaphor comprehension does not appear to be associated with people's background and expertise with spatial data, thus underlining the power of metaphorical mapping across user groups (see, for example, Fabrikant 2001a).

## Validating Spatializations

Spatializations combine intense computation with visual, interactive results. While a single spatialization product may involve diverse approaches, such as standard statistical inference, neural network models, and interactive visualization, these are associated with very different traditions for evaluating their validity. Significance tests, verification of trained neural networks, and human subject tests may address these approaches individually, but integrated evaluations are needed in order to ensure that a balance is achieved between the data relationships that exist in high-dimensional space and the patterns that can be communicated in a cognitively accessible, low-dimensional, representational space.

One direction of integrated validation potentially attractive to cartographers is to provide visual indicators of computational plausibility during the spatialization process, or as part of the final visual product geared towards analytical use. For example, one can embed into a spatialization visual cues regarding the stability of a clustering procedure. Figure 16 shows a combination of SOM neuron clustering with attributes of individual neurons and documents.



The “elevation” of individual neurons expresses the degree to which the three top-ranked terms dominate a neuron vector. The higher a neuron, the narrower is its topical focus. “Mountains” in the visualization correspond to well defined topics in the data set, while “pits” indicate a lack of focus. Clusters that incorporate extremely low-lying areas are to be viewed with particular suspicion. The landscape visualization helps to explain the apparent heterogeneity of some clusters, as expressed by mismatched cluster labels depicted in Figure 16. The cluster labeled “ethnic”—“production”—“new” is a prime example for this. The visualization also helps to demonstrate the cause of cluster heterogeneity, i.e., the pulling together of documents containing few index terms when using a Euclidean distance measure. With hierarchical clustering, this heterogeneity propagates as one moves to higher-level clusters. This is apparent with the cluster labeled “landscape”—“sediment”—“population” in the left portion of Figure 14.

## Summary and Outlook

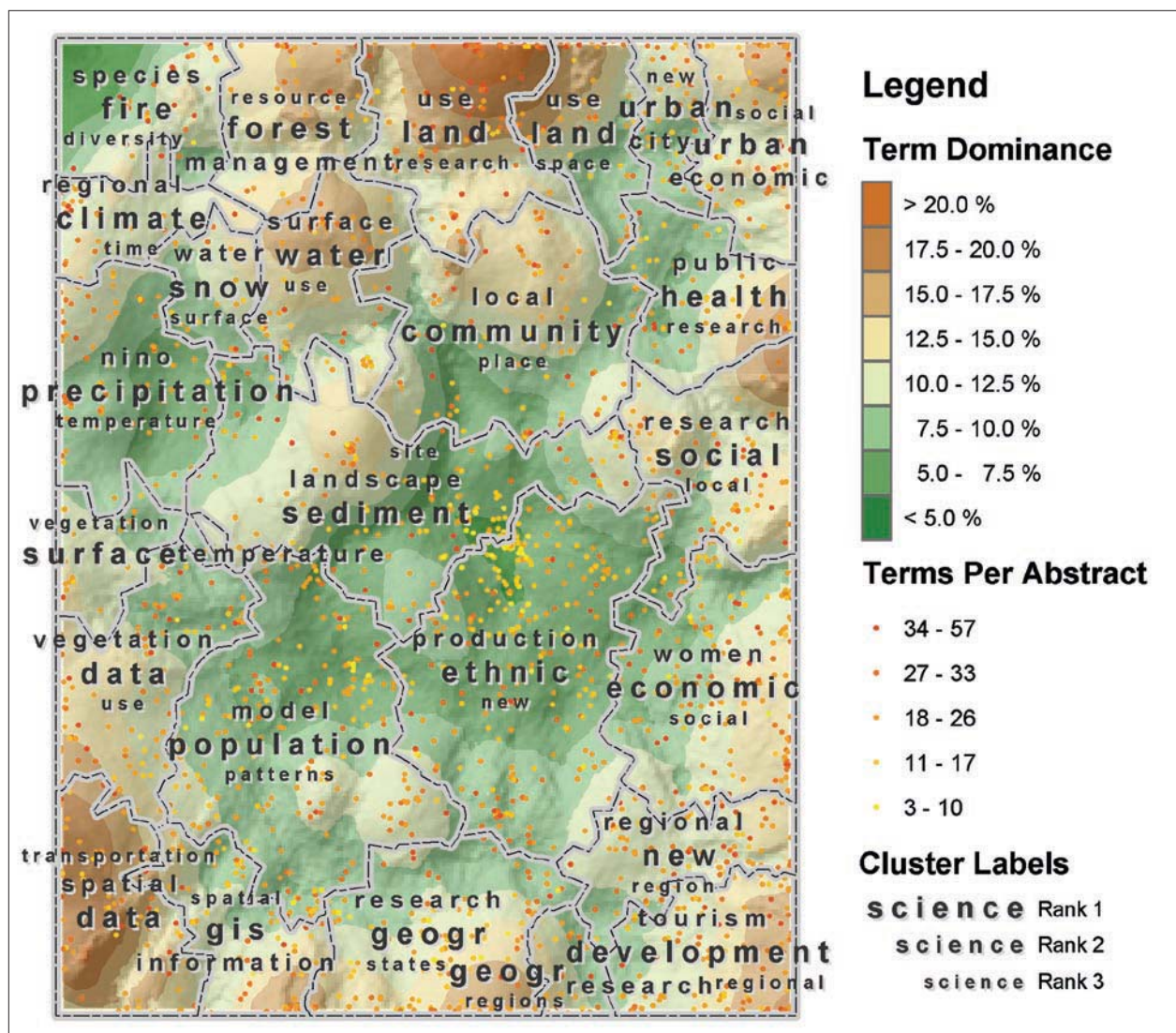
This paper has demonstrated that increasing involvement in non-geographic information visualization can provide cartography and GIScience with a unique opportunity to participate in the development of an evolving interdisciplinary endeavor. What emerges from the description of principal approaches, specific methods, and implementation examples is a research agenda that accommodates a broad spectrum of our discipline by posing a range of research challenges.

The first and foremost challenge is to acknowledge the relevance of cartographic expertise beyond the visualization of *geographic* phenomena. It is important to separate the multitude of cartographic and GIScience methods from the geographic reality to which these have traditionally been applied. This requires a more thorough understanding of the essence of cartographic activity. Since the rise of analytic cartography we have known that cartography is, more than anything else, about spatial data transformation (Tobler 1979). With current geovisualization approaches we now have the necessary methods in hand to provide knowledge transfer into related research communities with highly interactive tools based on solid semantic foundations (MacEachren 1995).

Increased awareness of, and empirical knowledge about, spatial cognition, including current work by GIScientists on geographic ontology, needs to be integrated into spatialization research. Current work at the University of California at Santa Barbara and

the University at Buffalo is combining an ontological approach informed by GIScience theory with an empirical, experimental methodology borrowed from cognitive science to assess the usability of the spatialization methods and transformation procedures discussed in this paper. The goal of this research is to deepen our understanding of how spatializations are perceived and understood, and to derive practical design guidelines from these insights. Once these ground rules for cognitively plausible visualizations are established, the next challenge will be to ensure that one employs plausible computational methods. For example, if empirical work demonstrates that a given visualization method enables users to more easily perceive certain structures in a data set, then it must be ensured that this is justified by the actual existence of such structures in high-dimensional space, as opposed to being an introduced artifact of the particular technique. This goal is surprisingly hard to reach. The main role for cartographically driven spatialization research will thus be to attempt the difficult balance of three competing aspects: (a) the need to discover and/or convey high-dimensional structures; (b) the need to determine the appropriate use of dimensionality-reducing techniques that always lead to a distortion of high-dimensional relationships; and (c) the need to employ visualization techniques that are in tune with our understanding of human cognition of the real world and geographic maps. It is in this context that cartographic research has to critically engage the various information visualization methods and systems that have been put forward. Another promising venture lies in investigating the degree to which proposed information visualization principles and techniques may be applicable to geographic data. Worthwhile research topics in this area range from specific visualization methods to typologies of such techniques.

A different aspect of an evolving research agenda relates to how cartographers and other GIScientists can actually make the impact of their contributions felt. The main thrust of information visualization research is happening outside of GIScience. Therefore, the transfer of cartographic expert knowledge is dependent on an intimate involvement in the relevant research communities. This refers particularly to the core information visualization community, which is currently dominated by computer scientists, even though it is still a young and evolving field (Card et al. 1999). Future engagement in such conference series as the IEEE Symposium on Information Visualization (InfoVis) or the annual Information Visualization Conference in London would be particularly worthwhile. With few exceptions (such as the geovisualization efforts at the Pennsylvania State University and the International Cartographic



**Figure 16.** Visual support for evaluating cluster validity. The visualization is based on a 60-by-80 neuron SOM. It shows individual point locations for several thousand AAG conference abstracts, the 25-cluster level of a hierarchical cluster solution, ranked cluster labels, and an indication of how much the highest-ranked terms dominate particular regions. Low term dominance may indicate a lack of sharply defined themes and therefore the existence of relatively heterogeneous clusters.

Association Commission on Visualization and Virtual Environments), cartographic researchers are mostly absent from these meetings and associated publications. Leaving aside core spatialization research, this lack of cartographic engagement is particularly poignant due to the growing number of geographic visualizations encountered within the confines of the information visualization community, and which are appearing in widely adopted textbooks on the subject (e.g., Spence 2001).

Spatialization efforts are also encountered in a heterogeneous collection of other fields. Information and library scientists have a particular appreciation for traditional cartographic products and are quite supportive of cartographic involvement, specifically in

the area of knowledge domain visualization (Börner et al. 2002). Relevant research and potential collaborators are also found in knowledge discovery and data mining, and in the somewhat distinct communities of data engineering and discovery science.

Finally, cartographic research on non-geographic information visualization should particularly benefit from the emerging interdisciplinary Semantic Web efforts. Today, input to the most impressive spatializations tends to be either derived from rigid, subjectively formed, and rather incomplete hierarchies (e.g., Open Directory Project), or by making equally subjective choices among an array of complex knowledge discovery tools and procedures. In the near future, the Semantic Web will assist with the integra-

tion of heterogeneous information. It will ease the exploitation and exploration of relationships between information elements, leading to more complex, yet ultimately more meaning-bearing information spaces. This has intriguing consequences for cartographic involvement, as it may lead to a renewed interest among non-cartographers in how our community has managed to not only represent the infinitely complex geographic reality within a limited display space, but also do it in a manner that enables people to recognize their world within it.

#### ACKNOWLEDGMENTS

André Skupin's work is supported by the Louisiana Board of Regents Support Fund, grant # LEQSF(2002-05)-RD-A-34. Sara Fabrikant's research is supported by the National Imagery and Mapping Agency (NMA-201-00-1-2005). Valuable suggestions and encouraging comments by the guest editor and three anonymous reviewers are gratefully acknowledged.

#### REFERENCES

- Bederson, B.B., J.D. Hollan, K. Perlin, J. Meyer, D. Bacon, and G. W. Furnas. 1996. Pad++: A zoomable graphical sketchpad for exploring alternate interface physics. *Journal of Visual Languages and Computing* 7: 3-31.
- Börner, K., C. Chen, and K. Boyack. 2002. Visualizing knowledge domains. In: B. Cronin (ed.), *Annual review of information science and technology*, vol. 37. Medford, New Jersey: Information Today, Inc./American Society for Information Science and Technology. pp. 179-255.
- Buckley, A.R., M. Gahegan, and K.C. Clarke. 2000. *Geographic visualization*. University Consortium for Geographic Information Science. [<http://www.ucgis.org/emerging/Geographicvisualization-edit.pdf>].
- Buttenfield, B.P., M. Gahegan, H.J. Miller, and M. Yuan. 2000. *Geospatial data mining and knowledge discovery*. University Consortium for Geographic Information Science. [<http://www.ucgis.org/emerging/gkd.pdf>].
- Card, S.K., J.D. Mackinlay, and B. Shneiderman. 1999. *Readings in information visualization: using vision to think*. The Morgan Kaufmann series in interactive technologies. San Francisco, California: Morgan Kaufmann Publishers.
- Castelli, V., L.D. Bergman, I. Kontoyannis, C.-S. Li, J.T. Robinson, and J.J. Turek. 1998. Progressive search and retrieval in large image archives. *IBM Journal of Research Development* 42(2): 253-67.
- Chalmers, M. 1996. A linear iteration time layout algorithm for visualising high-dimensional data. *Proceedings of IEEE Visualization '96*, San Francisco, California. pp. 127-132.
- Chang, S.-F., J.R. Smith, M. Beigi, and A. Benitez. 1997. Visual information retrieval from large distributed online repositories. *Communications of the ACM* 40(12): 63-71.
- Chen, C., and M. Czerwinski (eds). 2000. Special issue on empirical evaluation of information visualisations. *International Journal of Human-Computer Studies* 53(5).
- Chen, C., and R. J. Paul. 2001. Visualizing a knowledge domain's intellectual structure. *IEEE Computer* 34(3): 64-71.
- Chen, H., P. Hsu, R. Orwig, L. Hoopes, and J. Nunamaker. 1994. Automatic concept classification of text from electronic meetings. *Communications of the ACM* 37(10): 56-73.
- Clarke, K. C., and P.D. Teague. 2000. Representation of cartographic uncertainty using virtual environments. In: *Proceedings, Accuracy 2000, 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, July 12-14, Amsterdam, Netherlands. pp. 109-116.
- Couclelis, H. 1998. Worlds of information: The geographic metaphor in the visualization of complex information. *Cartography and Geographic Information Systems* 25(4): 209-20.
- Dodge, M., and R. Kitchin. 2001. *Atlas of cyberspace*. Harlow, England; New York, New York: Addison-Wesley.
- Eades, P. 1984. A heuristic for graph drawing. *Congressus Numerantium* 42: 149-60.
- Fabrikant, S.I. 2000. Spatialized browsing in large data archives. *Transactions in GIS* 4(1): 65-78.
- Fabrikant, S.I. 2001a. Evaluating the usability of the scale metaphor for querying semantic spaces. In: D.R. Montello (ed.), *Spatial information theory: Foundations of geographic information science (Lecture Notes in Computer Science 2205)*. Berlin, Germany: Springer-Verlag. pp. 156-72.
- Fabrikant, S.I. 2001b. Visualizing region and scale in information spaces. In: *Proceedings of 20th ICA/ACI International Cartographic Conference*, August 6-10, 2001, Beijing, China. pp. 2522-9.
- Fabrikant, S.I., and B. P. Buttenfield. 1997. Envisioning user access to a large data archive. In: *Proceedings of GIS/LIS '97*, Oct. 28-30, Cincinnati, Ohio. pp. 686-92.
- Fabrikant, S.I., and B. P. Buttenfield. 2001. Formalizing semantic spaces for information access. *Annals of the Association of American Geographers* 91(2): 263-80.
- Fabrikant, S.I., M. Ruocco, R. Middleton, D.R. Montello, and C. Jørgensen. 2002. The first law of cognitive geography: Distance and similarity in semantic space. *Proceedings of GIScience 2002*, Boulder, CO. pp. 31-33.
- Furnas, G., and B. Bederson. 1995. Space-scale diagrams: Understanding multiscale interfaces. In: *Proceedings of ACM Conf Human Factors in Computing Systems (CHI 95)*, May 1995, Denver, Colorado. pp. 234-41.
- Gahegan, M., M. Takatsuka, M. Wheeler, and F. Hardisty. 2002. Introducing GeoVISTA studio: An integrated suite of visualization and computational methods for exploration and knowledge construction in geography. *Computers, Environment and Urban Systems* 26: 267-92.
- Girardin, L. 1995. *Cyberspace geography visualization*. [<http://www.girardin.org/luc/cgv/report/>].
- Goodchild, M.F., and D.G. Janelle. 1988. Specialization in the structure and organization of geography. *Annals of the Association of American Geographers* 78(1): 1-28.
- Healy, G., and A. Jain. 1996. Retrieving multispectral satellite images using physics-based invariant representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8): 842-8.

- Johnson, B., and B. Shneiderman. 1991. Treemaps: A space-filling approach to the visualization of hierarchical information structures. In: *Proceedings of IEEE Visualization '91*, October 1991, San Diego, California. pp. 275-82.
- Kamada, T., and S. Kawai. 1989. An algorithm for drawing general undirected graphs. *Information Processing Letters* 31: 7-15.
- Kohonen, T. 1995. *Self-organizing maps*. Berlin, Germany: Springer-Verlag.
- Kohonen, T., S. Kaski, K. Lagus, J. Salojärvi, T. Honkela, V. Paatero, and A. Saarela. 1999. Self organization of a massive text document collection. In: E. Oja and S. Kaski (eds), *Kohonen maps*. Amsterdam, Netherlands: Elsevier. pp. 171-82.
- Kruskal, J.B., and M. Wish. 1978. Multidimensional scaling. *Sage University Paper series on Quantitative Applications in the Social Sciences*, series no. 07-011. Beverly Hills, California: Sage Publications.
- Lamping, J., and R. Rao. 1996. The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of Visual Languages and Computing* 7(1): 33-55.
- Legrady, G., and T. Honkela. 2002. Pockets full of memories: An interactive museum installation. *Visual Communication* 1(2): 163-70.
- Leung, Y.K., and M.D. Apperley. 1994. A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction* 1(2): 126-60.
- Li, B. 1998. Exploring spatial patterns with self-organizing maps. In: *Proceedings of GIS/LIS '98*, Fort Worth, Texas. CD-ROM.
- Lin, X. 1992. Visualization for the document space. In: *Proceedings of IEEE Visualization '92*, October 1992, Boston, Massachusetts. pp. 274-281.
- Lloyd, R. 2000. Self-organized cognitive maps. *Professional Geographer* 52(3): 517-31.
- Ma, W.Y., and B.S. Manjunath. 1996. *A pattern thesaurus for browsing large aerial photographs*. ECE Technical Report 96-10, Department of Electrical and Computer Engineering, Santa Barbara, California.
- MacEachren, A.M. 1995. *How maps work*. New York, New York: The Guilford Press.
- MacEachren, A.M., and M.-J. Kraak (eds). 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science* 28(1).
- Manjunath, B.S., and W. Y. Ma. 1996. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8): 842-8.
- Mark, D.M., A. Skupin, and B. Smith. 2001. Features, objects, and other things: Ontological distinctions in the geographic domain. In: D.R. Montello (ed.), *Spatial information theory: Foundations of geographic information science (Lecture Notes in Computer Science 2205)*. Berlin, Germany: Springer-Verlag. pp. 488-502.
- Miller, H.J., and J. Han. 2001. *Geographic data mining and knowledge discovery*. Research monographs in geographic information systems. London, U.K.; New York, New York: Taylor & Francis.
- Mulcahy, K.A., and K.C. Clarke. 2001. Symbolization of map projection distortion: A review. *Cartography and Geographic Information Science* 28(3): 167-81.
- Ormsby, D., and W. Mackaness. 1999. The development of phenomenological generalization within an object-oriented paradigm. *Cartography and Geographic Information Science* 26(1): 70-80.
- Salton, G. 1968. *Automatic information organization and retrieval*. New York, New York: McGraw-Hill.
- Sammon, J.W. 1969. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* C-18: 401-9.
- Sarkar, M., and M.H. Brown. 1994. Graphical fisheye views. *Communications of the ACM* 37(12): 73-84.
- Schvaneveldt, R.W. 1990. *Pathfinder associative networks: Studies in knowledge organizations*. Norwood, New Jersey: Ablex.
- Sheikholeslami, G., A. Zhang, and L. Bian. 1999. A multi-resolution content-based retrieval approach for geographic images. *GeoInformatica* 3(2): 109-39.
- Skupin, A. 1998. *Organizing and visualizing hypermedia information spaces*. Ph.D. thesis, State University of New York at Buffalo, Buffalo.
- Skupin, A. 2000. From metaphor to method: Cartographic perspectives on information visualization. In: *Proceedings of InfoVis 2000*, October 2000, Salt Lake City, Utah. pp. 91-7.
- Skupin, A. 2002a. A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications* 22(1): 50-8.
- Skupin, A. 2002b. On geometry and transformation in map-like information visualization. In: K. Börner and C. Chen (eds), *Visual interfaces to digital libraries (Lecture Notes in Computer Science 2539)*. Berlin, Germany: Springer-Verlag. pp. 161-70.
- Skupin, A., and B. P. Buttenfield. 1996. Spatial metaphors for visualizing very large data archives. In: *Proceedings of GIS/LIS '96 Annual Conference and Exposition*, November 19-21, Denver, Colorado. pp. 607-17.
- Skupin, A., and B.P. Buttenfield. 1997. Spatial metaphors for visualizing information spaces. *Proceedings of ACSM/ASPRS Annual Convention and Exhibition*, April 19-21, Seattle, Washington. pp. 116-25.
- Spence, R. 2001. *Information visualization*. Boston, Massachusetts: Addison Wesley.
- Suciu, D. 1998. An overview of semistructured data. *SIGACT News* 29(4): 28-38.
- Tenenbaum, J.B., V. de Silva, and J.C. Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290: 2319-23.
- Tobler, W.R. 1970. A computer model simulating urban growth in the Detroit region. *Economic Geography* 46(2): 234-40.
- Tobler, W.R. 1973. A continuous transformation useful for districting. *Annals of the New York Academy of Sciences* 219(9): 215-20.
- Tobler, W.R. 1976. The geometry of mental maps. In: R.G. Golledge and G. Rushton (eds), *Spatial choice and spatial behavior*. Columbus, Ohio: The Ohio State University Press. pp. 69-82.
- Tobler, W.R. 1979. A transformational view of cartography. *The American Cartographer*, 6(2): 101-106.

- Tobler, W.R., and S. Wineberg. 1971. A Cappadocian speculation. *Nature* (231): 39-41.
- Töpfer, F., and W. Pillewizer. 1966. The principles of selection. *Cartographic Journal* 3(1): 10-16.
- Ultsch, A. 1993. Self-organizing neural networks for visualization and classification. In: O. Opitz, B. Lausen, and R. Klar (eds), *Information and classification*. Berlin, Germany: Springer-Verlag. pp. 307-13.
- Van Der Wel, F.J.M., R. M. Hootsmans, and F. Ormeling. 1994. Visualization of data quality. In: A.M. MacEachren and D.R. Fraser Taylor (eds), *Visualization in modern cartography*. New York, New York: Elsevier Science. pp. 313-31.
- Wise, J.A. 1999. The ecological approach to text visualization. *Journal of the American Society for Information Science* 50(13): 1224-33.
- Wise, J.A., J.J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. 1995. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In: *Proceedings of InfoVis '95*, Atlanta, Georgia. pp. 51-8.
- Zhang, J.X., and M.F. Goodchild. 2002. *Uncertainty in geographical information*. New York, New York: Taylor and Francis. 