Visual Analysis of Spatiotemporal Multilevel Data

René Rosenbaum, Hans-Jörg Schulz, Steffen Hadlak, and Heidrun Schumann

Abstract— For multilevel data, levels are not the result of a hierarchical aggregation, but contain independently produced data. While first visualization techniques for this kind of data exist, suitable interactive exploration techniques have rarely been investigated so far. In this paper, we introduce means for the analysis, representation, and exploration of spatiotemporal multilevel data as first solutions to fill this gap. Particularly, these are analysis strategies serving to order the data appropriately, a novel data display providing an overview to the data as well as level- and path-based exploration strategies supporting systematic data browsing. Appropriate level and path filtering techniques are also provided. Results from an application of this technology to election data indicate their usefulness for conducting visual analysis tasks in spatiotemporal multilevel data.

1 Introduction

Interactive analysis and exploration is a necessary approach when foraging for information in an unknown complex data set, which is too large to be shown in all its entirety and all its detail. One such complex kind of data is *multilevel data* [9], whose complexity originates from being structured in a set of levels. These levels are not necessarily ordered, but might be arranged in a hierarchal fashion to express organizational structures in the data. However, there might be complex inter- and intra-level relationships as each level can hold a different kind of data. Thus, multilevel data poses a challenge to the traditionally accepted way of exploring data, as the lack of a sensible overview prohibits an Overview&Detail-style exploration via drill-down in data space or zoom-in for the view space. Drill-down operations on aggregated data bring forth more detailed information, which are usually predictable from the aggregated overview [2]. Yet for multilevel data, they produce entirely new information without giving a hint of what may lie on levels beneath and whether it is of any interest for the current goal of analysis. Due to this, multilevel data must always be explored in its entirety to grasp them.

This paper introduces a novel concept for the interactive analysis and exploration of multilevel data in a visual manner. Central to our approach is a novel data representation, the level view, which is designed to provide an overview of the different levels within the data. As we limit ourselves to spatiotemporal data, we apply the Triad Representational Framework [5] to meaningfully organize a large number of levels. A number of secondary data-dependent data displays support gaining more detail to levels of current interest to the user. Despite their complex structure, we support the exhaustive and systematic exploration of multilevel data by pinning and filtering technology. The whole visual analysis process consists of the following three main components:

1. **Analysis** - serving to structure and order the data.
2. **Visual representation** - providing a novel visualization of the data levels and their relationships as well as two data specific displays.
3. **Exploration** - serving to investigate data levels and their relationships by novel means for browsing and filtering.

The contribution of this publication can be summarized as follows:

- A concept for the interactive visual analysis of spatiotemporal multilevel data
- A novel data view for the visual representation of multilevel data
- Two novel strategies to interactively explore large multilevel data sets

- *All authors are with the University of Rostock. e-mail: {rrosen,hjschulz,hadlak,schumann}@informatik.uni-rostock.de*

In Section 2, the main properties and characteristics of multilevel data are described. Section 3 reviews related research. The proposed concept is introduced in Section 4 and its applicability shown by a prominent use case in Section 5. Section 6 closes with conclusions and remarks for future research.

2 Multilevel data

Along the lines of the data model presented in [4], the multilevel property defines a multipartite information structure on top of an arbitrary number of data items, labeled *information objects* (IOs). Each individual *level* may stem from a different source and thus usually represents a different kind of IO. A single level consists of one type of IO only. There is usually no explicit *inter*- or *intra-level* order. However, some levels can be arranged hierarchically. This hierarchy can be implicitly given for levels containing related IOs, such as seconds, minutes, and hours define three hierarchical levels of temporal data, but is often conceptual. In the latter case, the order must be made available from external sources. While implicit hierarchies can usually be aggregated, conceptual hierarchies cannot.

Multilevel data can be found in many application fields. One prominent example from the field of biology is the human body. It can be decomposed in multiple levels, such as organ systems, organs, cells, and gene expressions. The containment property thereby defines a conceptual hierarchy on the individual parts, e.g., organs are placed on a higher level than cells. Due to their specifics, however, parameters and functions of none of these levels can be inferred simply by aggregation.

Another example are spatiotemporal election data. Such data are assessed in different manners, such as for whole countries or local counties, for whole years or individual days, for all or individual parties, for all or individual associated candidates, with or without participation rates, and so forth. Each assessment leads to a certain number of IOs and is assigned an individual level. Levels representing solely spatial or temporal data can be merged taking advantage of their inherent hierarchy. The levels “party” and “candidates” may also be arranged hierarchically due to the fact that individual or multiple candidates are part of the larger group of party members. In this case, external knowledge is used to define this order.

3 Related Work

Since multilevel data is often equated with hierarchically aggregated data, a glance over prior work on multilevel visualization and exploration conveys the impression that a plethora of suitable methods already exist. Almost all of these techniques consider data on a single, lowest level, which is abstracted or aggregated into higher levels. In doing so, these methods rely on the fact that higher levels naturally subsume the data on lower levels.

One example, in which this becomes apparent is the data cube, as it is used in Online Analytical Processing (OLAP) [8, 1]. Such a data cube is composed of cells, which contain the individual data values on a lowest level of granularity. Higher-level views are then generated

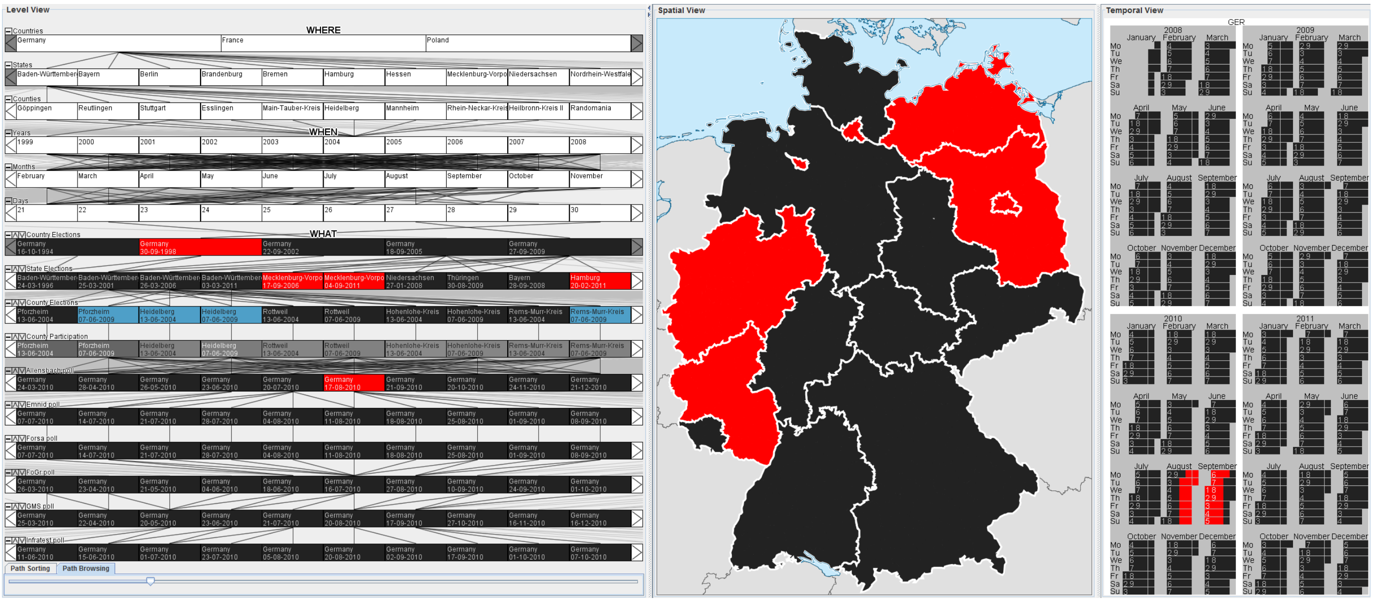

Fig. 1. The display setup introduced to visually explore spatiotemporal multilevel data: The level view (left) is designed to visualize the distinct characteristics of multilevel data. Single levels are represented as horizontal bars consisting of individual information objects. Color-coding and labeling is applied to easily convey their most important attributes (the most popular party in this case). As certain levels define a substructure, these can be preserved by partitioning the levels in semantic groups (Where, When, What?). Top-down paths connect associated information objects residing at different levels and form the basis for a structured path-based browsing. An additional map (center) and calendar display (right) provide intuitive representations for a selected level in space and time and also serve as a means for level-based exploration.

by aggregating all individual cells within a certain subvolume of the cube. These data cubes are the embodiment of the mechanistic idea that the sum of the parts equals the whole. Yet, this property does not hold, for example, for conceptual hierarchies. In this case, each level is described individually by its own data, which can no longer be represented as a data cube and OLAP cannot be used to query them.

For data without such an aggregation hierarchy, approaches for completely unrelated levels in the sense of individual data dimensions are probably the closest fitting visualization and query methods that exist so far. An example is the Jigsaw system [7], which arranges dimensions/levels in columns and draws polylines across them where data tuples combine their values. Many other visualization techniques follow a similar principle, such as Parallel Sets [3] or Sankey Diagrams [6]. While these techniques no longer rely on the strong hierarchy building through aggregation, they introduce a new restriction instead: the co-occurrence in a data tuple. This is again not in tune with the multi-level data paradigm, which generates independent data for each of the levels.

While the latter approach is close enough to serve as a first representation for multilevel data, its interactive exploration is a likewise open field of research. Since established exploration patterns cannot be employed for multilevel data, the following describes our new visual analysis approach for such data.

4 A Concept for the Visual Analysis of Spatiotemporal Multilevel Data

This section introduces a novel concept for the visual analysis of multilevel data. Although it is designed to be applied to arbitrary data sources of this kind, it is described from the perspective of spatiotemporal data. The main goal in the visual analysis of multilevel data is the discovery of their intra- and inter-level relationships. In order to accomplish this, we introduce novel means for their *analysis*, *visual representation* as well as *exploration*.

4.1 Analysis

Multilevel data are inherently structured in sets of levels and associated IOs. The analysis component provides means to order these sets introducing further structure. *Inter-level ordering* is used to group and sort on a level basis. *Intra-level ordering* serves to arrange the individual IOs of a particular level.

4.1.1 Inter-level ordering

Due to the fact that there are many different levels, it is meaningful to provide appropriate level groupings and orderings as a foundation for their appropriate visual representation and exploration. *Grouping* mainly serves to arrange related levels in close proximity. *Ordering* sorts the levels dependent on relevance. Both kinds of rearrangements are data-dependent, but are also determined through the analysis requirements of the application domain in whose context the data are explored. As one example in the context of spatiotemporal data, we support the grouping of levels containing IOs related to space and time. They are collected, separated, and ordered with regards to their underlying hierarchy. All levels are then grouped using the “Where, When, What?” paradigm and can now be accessed by spatial, temporal, and general aspects individually.

4.1.2 Intra-level ordering

In order to overcome the problems imposed by a large number of IOs, we support intra-level ordering ensuring that they appear in an order most appropriate for the user and given goals. One meaningful result is the placement of most interesting IOs at the beginning of each level where they are immediately accessible. To consider existing inter-level relationships between IOs, intra-level ordering may also place related IOs residing at different levels at similar positions within each level. Due to the fact that each level represents a different kind of data, however, sorting is usually performed independently for each level. Different intra-level orderings are possible. Temporal IOs are sorted in ascending or descending order. For the ordering of spatial IOs we employ a specific geographic sorting that orders the data from north-west to south-east. Ordering of other data levels depends on the respective kind of IO and range from simple succession to user- or goal-dependent criteria, such as pre-defined ranking.

4.2 Visual Representation

Multilevel data cannot be meaningfully summarized or aggregated and thus must be shown in their entirety. In order to support the overview-to-detail principle in their visual representation, we introduce a novel data display, the *level view*, providing a notion to all inherent levels and IOs at a glance. To gain further detail to the spatiotemporal properties of a selected level, we use secondary *data displays* representing the respective IOs in an individual and more common form.

An example of a level view is shown in Figure 1, left. Each level is represented as a horizontal bar. All levels are aligned in a top-to-bottom fashion. IOs associated to a level are visually encoded by individual color-coded sections. The color-coding is data-driven and thus depends on the respective level IOs. To establish connection between neighboring levels, associated IOs are visually linked. To support the prominent display of most important data, we take advantage of the results from the analysis stage; levels and IOs ordered so that most relevant items are placed on the top/left position of the view and grouped to support quick access to the main characteristics of spatiotemporal data.

One of the benefits of the level view is its ability to show levels and IOs in their entirety. Inter- and intra-level relationships can be conveyed without changing the view. This is of strong advantage compared to related approaches as this allows a user to compare and relate many different aspects and properties of the data at a glance.

In order to provide appropriate representations for the IOs of a particular level, additional data displays are supported (see Figure 1, center and right). In the context of spatiotemporal data, we provide a map display to consider the specifics of spatial IOs and a calendar display to better convey temporal aspects of the data.

4.3 Exploration

The exploration stage allows the user to rearrange and browse the different levels, their IOs, and connections dependent on current interests and goals. Since the nature of multilevel data does not permit the common refinement approach of a drill-down exploration, other systematic ways of exploring such data have to be found. We propose to take advantage of principles well-known in the field of graph analysis: *level-based exploration*, which allows the user to traverse the data level by level in a breadth-first search order, and *path-based exploration*, which passes through the data by going from one top-to-bottom path to the next in a depth-first search order. The option of a quick back and forth between both kinds of exploration enables effortless adaptation to shifts of the analysis goal. We also support means to *filter* the large number of levels and IOs dependent on current needs.

4.3.1 Level-based Exploration

Level-based exploration considers a single data level at once. We support this by the level view as well as the map and calendar display.

The level view well supports level-based exploration as all IOs associated to a particular level are arranged horizontally and thus are visually separated. However, the initial intra-level ordering resulting from the analysis stage is often not enough to reveal all interesting insights of the data. In order to support the required adaptation to current needs, we provide means to specify the desired sorting criteria interactively.

A systematic level-wise exploration can be achieved by subsequently browsing through all relevant levels in a top-to-bottom fashion. In order to underpin the importance of time and space in spatiotemporal data, level-based exploration is further supported by the map and calendar display. Within those displays, the IOs of a selected level are encoded in their "natural" spatial or temporal context (see Figure 1, center and right). Especially for spatial data, this allows for a much better representation of locations compared to the level view. However, only one level can be represented at a glance. This leads to an increased number of interactive level switches and in turn to more cognitive strain compared to the level view showing many different levels at once.

4.3.2 Path-based Exploration

Path-based exploration is most suitable when inter-level relations are of interest. Due to the potentially large number of levels and complex IOs, this can only be provided by the level view. Here, inter-level associations are indicated by an entire top-to-bottom level path connecting associated IOs. Although, inter- and intra-level ordering as a result of automated analysis can often provide an appropriate initial view to the data, exploratory tools are required for in-depth analysis. We provide interactive means for *linking* and *ordering*.

There are usually multiple options to link the IOs residing at neighboring levels. We support joining levels based on their spatial or temporal attributes. To specify, whether space or time is used for linking, we take advantage of the Point-of-Interest paradigm. By selecting a specific time- or space-related IO within the "When"- or "Where"-group respectively, the linking is adapted to the chosen aspect.

We also support to interactively place levels that promise to show interesting relationships close to each other significantly reducing the efforts required to trace many individual paths instead. To prevent strong horizontal scattering of a path, we take advantage of prior analysis placing associated IOs in close proximity. A structured path-based exploration is then achieved by browsing through the individual paths, typically from left to right i.e. important to less important.

4.3.3 Filtering

Large and complex information spaces clutter the level view. To overcome this, we support means for *level filtering* and *data pinning*.

Level filtering can be accomplished by the automated or interactive removal of individual or groups of levels. It is strongly demand-driven. To convey the existence of filtered levels they are represented by placeholders (see Figure 2, "What"-group). In order to reduce the potentially very large set of paths to explore, we take advantage of the concept of *pinning*. Individual IOs can be interactively selected in one or multiple levels and the available paths will be cut down to those related to the pinned IOs. This further supports the Point-of-Interest paradigm and permits to investigate the dependencies between the different levels. It also pinpoints paths of interest in a more local fashion than the ordering allows. Due to the fact that all provided data views are coordinated, pinning is also supported by the map and calendar

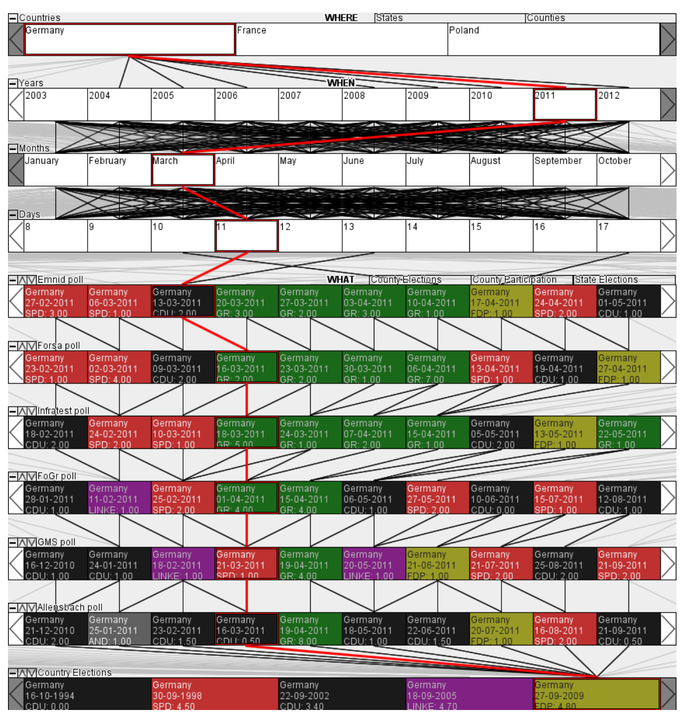

Fig. 2. Effect of the March 11, 2011 Fukushima Daiichi nuclear disaster on German voter behavior. The figure clearly indicates a strong increase in the popularity of the German Green party (green) compared to other parties (black, gray, red, yellow, violet) right after the event (red path).

display in order to allow for a more intuitive selection of interesting locations or points in time.

The next section will detail results of a case study we performed on German election data using the introduced means for visual analysis.

5 Case Study

This section presents a case study that is based on German election data compiled from different sources and time periods. This truly multilevel data consist of 16 levels structured in three hierarchical time levels (years, months, days), three hierarchical space levels (countries, states, counties), and ten data levels covering different results and aspects of elections (see to Section 2). This use case has been selected out of the many possible applications fields of the introduced concept due to its spatiotemporal background.

The initial level view shows all levels to allow for an overview to the available IOs and connections between them. All levels are ordered and linked by time to discover tendencies in election behavior. Map and calendar display indicate location and time of the IOs associated to the first level of the "What"-group, a voter poll in this case. The user can now start exploring using the described means for level- and path-based exploration as well as level filtering and pinning.

While exploring the collected election data, we revealed two interesting events and their consequences on German voter behavior that would be cumbersome to find without our approach: (1) the Fukushima Daiichi nuclear disaster (temporal effects) and (2) the Stuttgart21 railway project controversy (temporal and spatial effects).

Figure 2 shows the level view, where the levels represent the winners with regard to the highest increase in voter popularity. The bottom level represents the official election data (time period of five years), the other colored levels monthly or weekly voter polls collected by different agencies. Path-based exploration and pinning of the particular date revealed that shortly after the Fukushima Daiichi nuclear disaster, the Green party had the largest increase in voter popularity across all parties, which was most certainly due to an increased support for their stance against nuclear power. This continued for over a month until the impact of the event became weaker. This might be due to the fact that the German government later decided to abandon nuclear power – a decision, which was not shared by all voters.

Figure 3 presents another finding from the data set and involves spatial and temporal effects. Stuttgart21 is a controversial railway and urban development program in the German state Baden Württemberg. It sparked debate and grassroots protest movements all across the country. The local state election held in the peak of the protest reflects the strong discontent of the local population with this project. Taking advantage of level-based exploration using the level holding the respective increase in voter popularity for all parties, the map clearly reveals that the Green party and numerous oppositional voters' associations were the winners of this election. This effect is very strong around the affected region and decreases with distance. The map display, however, can capture a single point in time only. The temporal changes can be explored more thoroughly in the level view. Applying level-based exploration on the first level in the "What"-section representing the official election data available for Stuttgart, it can be seen that there was a strong increase in the popularity of the Green party suddenly after a long period dominated by the party traditionally governing the state. By taking advantage of pinning it can be seen that this happened right after the announcement of the project. By taking advantage of path-based exploration, it was also revealed that despite the strong increase, the alternative parties were not able to overtake the government (second level of "What"-section) and that it seems as if the results of previous federal elections (third level of "What"-section) did not affect local voter behavior. Such discoveries based on a multitude of levels are very difficult to make using existing approaches.

6 Conclusion

We introduced a novel concept for the visual analysis of spatiotemporal multilevel data. We proposed means to analyze the data transferring them into a more structured order. Furthermore, the proposed level view provides an immediate overview to the data, while the secondary calendar and map displays facilitate a more detailed examination of individual levels in the context of time and space. We also introduced different mean for interaction facilitating level- and path-based data exploration and the reduction of the displayed data to an appropriate level. As shown by a case study this can lead to deep insight without requiring much exploratory efforts.

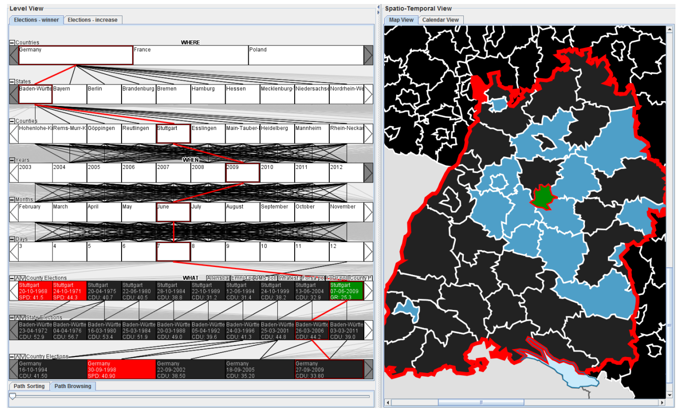

Fig. 3. Effect of a controversial governmental project in Stuttgart on local voter behavior. The map clearly shows the local effects of this event. The Green party (green) had the strongest increase right after the event followed by different voters' associations (blue). As shown in the level view (first level of "What"-section), this event broke a long trend in local voter behavior, but does not seem to have any influence on the state and federal elections (second and third level of "What"-section).

This paper represents just a first step towards a systematic view to the visual analysis of multilevel data. In future work, we are concerned with two primary research directions: (1) evaluation and (2) technical improvement. A preliminary user study already revealed that level- and path-based exploration are applied by the users with quite distinct objectives. In a profound user study, we will examine the properties of each strategy in more detail leading to statements when to use which method in a given application context. Further, we will develop different ordering techniques (analysis) to deal with complex IOs and a novel path representation (visual representation) taking advantage of edge bundling to provide overviews to common routes.

References

[1] V. Beaulieu and Y. Bedard. Interactive exploration of multi-granularity spatial and temporal datacubes: Providing computer-assisted geovisualization support. In *GeoVA'08: Proceedings of the Workshop on Geospatial Visual Analytics*, 2008.

[2] N. Elmqvist and J.-D. Fekete. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, 16(3):439–454, 2010.

[3] R. Kosara, F. Bendix, and H. Hauser. Parallel Sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):558–568, July-August 2006.

[4] M. Kreuseler and H. Schumann. A flexible approach for visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):39–51, 2002.

[5] D. Peuquet. *Representations of Space and Time*. The Guilford Press, 2002.

[6] P. Riehmann, M. Hanfler, and B. Froehlich. Interactive Sankey diagrams. In *InfoVis'05: Proceedings of the IEEE Symposium on Information Visualization*, pages 233–240, 2005.

[7] J. Stasko, C. Görg, and Z. Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[8] C. Stolte, D. Tang, and P. Hanrahan. Multiscale visualization using data cubes. *IEEE Transactions on Visualization and Computer Graphics*, 9(2):176–187, 2003.

[9] M. Streit, H.-J. Schulz, D. Schmalstieg, and H. Schumann. Towards multiuser multi-level interaction. In *CoVIS'09: Proceedings of the Workshop on Collaborative Visualization on Interactive Surfaces*, pages 5–8, 2009.