TOWARD THE ROLE OF INTERACTION IN VISUAL ANALYTICS

Andreas Kerren

Linnaeus University
School of Computer Science, Physics and Mathematics
Vejdes Plats 7
SE-351 95 Växjö, SWEDEN

Falk Schreiber

Martin Luther University Halle-Wittenberg
Von-Seckendorff-Platz 1
D-06120 Halle, GERMANY, and
IPK Gatersleben
Corrensstrasse 3
D-06466 Gatersleben, GERMANY

ABSTRACT

This paper firstly provides a general introduction in the most important aspects and ideas of Visual Analytics. This multidisciplinary field focuses on the analytical reasoning of typically large and complex (often heterogeneous) data sets and combines techniques from interactive visualizations with computational analysis methods. Hereby, intuitive and efficient user interactions are a fundamental component which has to be efficiently supported by any Visual Analytics system. This integration of interaction techniques into both visual representations and automatic analysis methods supports the human-information discourse and can be realized in various ways which is discussed in the second part of the paper. We give examples of possible applications of Visual Analytics from the domain of biological simulations and highlight the importance and role of the human in the analysis loop.

1 INTRODUCTION

For many years, people from various science and business domains have been overwhelmed by huge and complex data collections that have to be analyzed. Local data storage is very cheap, and data sets are more and more moved into the Cloud where either individuals or collaborators have access to them. According to the HMI Report (Short et al. 2011), 9.57 Zettabyte (= $9,57 \cdot 10^{21}$) of new information was processed on enterprise servers in 2008 worldwide (for comparison: 5 Exabyte (= $5 \cdot 10^{18}$) in 2002). Thus, for instance, 63 Terabytes of information were processed by one company on average in that year. Such figures let assume that this process of data collection and storage will heavily increase in the future. This is also true in the sciences where, for example, the massive collection of sensor data in the environmental sciences (Bryant et al. 2008) or the huge amount of *omics data in systems biology automatically generated by high-throughput technologies (Gehlenborg et al. 2010) lead to the challenge of interpreting all of these data sets. The fundamental problem today is to transform the data—which is typically not pre-processed, erratic, stored in idiosyncratic formats, sometimes uncertain, and often composed of various types (multidimensional, time-dependent, geo-spatial, . . .)—into information and make it useful/available/analyzable to analysts. Often, this challenge is called the *information overload problem* (IOP). Positive effects of such a transformation are then to discover something that is interesting (like patterns or outliers), to make decisions in crisis situations, or to monitor a huge data set in real time and under time pressure.

How can Visual Analytics (VA) help to solve the information overload problem? "Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2006). A crucial property of this field is that computational methods of data analysis are combined with interactive visualization techniques in order to analyze data more efficiently. For many data analysis problems, fully automated analysis methods only work for well-defined and well-understood problems, i.e., there has to exist a model of the underlying problem (Keim et al. 2008). Otherwise, traditional data mining techniques

will not work. Even if a model exists, then the results of the automated analyses have to be sufficiently communicated to and interpreted by analysts. Here, interactive visualizations come into the play as they are able to support the analyst to discover (possibly unexpected) patterns, trends, or relationships in the data. Therefore, computational methods for the analysis of huge, complex data sets in combination with an interactive visual analysis are indispensable to capture important correlations or patterns that will otherwise go unnoticed. (Ceglar et al. 2003) stated that "user participation in the [data] mining process results in greater confidence in the correctness of the discovered patterns due to the sense of control that guidance capabilities provide". Interaction allows, among other things, to explore "unknown" data collections following Shneiderman's mantra of information visualization *overview first, zoom and filter, details on demand* (Shneiderman 1996) or to build hypotheses with the help of "What if?"-questions and to verify them visually or with algorithmic methods. This important insight is reflected by the head note of Visual Analytics: "*detect the expected and discover the unexpected* from massive and dynamic information streams and databases consisting of data of multiple types and from multiple sources, even though the data are often conflicting and incomplete" (Thomas and Cook 2006). To sum up, the need to combine interactive visualization with computational analysis methods is obvious and opens novel possibilities to address the IOP. "The challenge is to identify the best automated algorithm for the analysis task at hand, identify its limits which can not be further automated, and then develop a tightly integrated solution with adequately integrates the best automated analysis algorithms with appropriate visualization and interaction techniques." (Keim et al. 2008).

This paper provides an overview of Visual Analytics with a special focus on the importance of the human in the analysis loop as well as on various types of user interaction which are fundamental for supporting the analysis process. We give examples of VA approaches from the application domain of systems biology because of several reasons: first, the available space in this paper is too short to introduce challenges of several application domains. Second, systems biology contains a set of highly-interesting grand challenges that is based on the analysis of vast, high-dimensional and complex data sets. Third, biological data have to be interactively explored as researchers often do not know in advance what they are looking for. Last, modeling and simulation plays an important role in biology, and there are biological data sets that are based on simulations, such as metabolic simulations, which might also be of interest for the simulation community.

The remainder of the paper is organized as follows. In Section 2, we briefly give an overview of important literature and initiatives in the field of Visual Analytics. Section 3 discusses different facets of Visual Analytics in more detail and highlights the role of interaction in the visual analysis process. Then, some example applications from the field of biology are highlighted in Section 4. Here, sample simulation methods are shortly presented, and the role of the human analyst and interaction possibilities during reconstruction and investigation of metabolism are covered. Finally, Section 5 concludes this paper.

2 BOOKS AND INITIATIVES ON VISUAL ANALYTICS

One of the first occurrences where the term "Visual Analytics" appeared is the introduction of a special issue of IEEE Computer Graphics and Applications written by (Wong and Thomas 2004) and published at autumn 2004. Shortly after this journal issue, a book on the research and development agenda of VA was published, called *Illuminating the Path* (Thomas and Cook 2005). It was the result of a crosscutting panel of research leaders in several fields assembled by the National Visualization and Analytics Center (NVAC) in the United States. The agenda is a comprehensive overview of the main idea of VA, needs and grand challenges. But certainly, some characteristics and concepts of VA popped up already earlier in the shape of specific methods and tools. Several excerpts and summaries of Illuminating the Path were published subsequently, such as (Thomas and Cook 2006).

In 2007, a seminar on Information Visualization took place at the International Conference and Research Center for Computer Science (Dagstuhl Castle) in Germany. The seminar discussions also included topics of VA, and the findings were released as book contribution that explains the difference between VA and

InfoVis (Keim et al. 2008). It identifies the technical challenges faced by VA researchers and describes a number of typical applications.

During the following years, several initiatives were started in order to further develop VA techniques and to build an own community. One of these ventures was *VisMaster* (VisMaster 2012), the European Coordination Action Project on Visual Analytics. As outcome of this two-year project, a book was compiled that describes a roadmap of VA research (Keim et al. 2010). In contrast to the previously mentioned books, it illuminates strategies for future research on the basis of an analysis of the state-of-the-art and open challenges and also exemplifies these ideas with many screenshots of tools in action. For an overview of the most important research challenges in VA, we refer the interested reader to this book. An accompanying video was also produced to demonstrate the importance of VA for Europe (VisMaster Video 2012), and a new European Web portal on VA (Visual-Analytics.EU 2012) was finally created. The international Web portal on VA is called *VA Community* (VA Community 2012).

There are several conferences and workshops on VA, but most of todays visualization conferences invite papers on this field too. The IEEE Conference on Visual Analytics Science and Technology (IEEE VAST) was founded in 2006 (originally as IEEE Symposium on Visual Analytics Science and Technology) and is the first international conference dedicated to advances in this field. Noteworthy is the co-located VAST challenge/contest: research groups are invited in advance to develop concrete solutions for given large and complex data sets and specific tasks that should be supported by the tools. Professional analysts then interact with the tools and provide feedback during the conference. A smaller event is, for example, the annual EuroVA workshop which is organized under the umbrella of the EuroVis conference.

3 FACETS OF VISUAL ANALYTICS

As already discussed in the introduction, VA is an integral approach for addressing the information overload problem that combines several disciplines, such as visualization, automatic data analysis and human factors. The last mentioned term covers fields that are important for ensuring human-centered aspects of VA systems. We summarize those related research fields together with the most important features in the following list which is mainly based on the work (Keim et al. 2008):

- **Interactive Visualization:** This field is often defined as the use of computer-supported, interactive, visual representations of data to amplify cognition (Card et al. 1999). Here, interaction techniques (e. g., focus&context, filtering, zooming&panning, . . .) are of particular importance to visually analyze large volumes of data. Information Visualization mainly focuses on the visualization of abstract data, whereas Scientific Visualization primarily aims at spatial data sets. There are a lot of subfields which cover specific application domains, such as BioVis, SoftVis, GraphVis, etc.
- **Automatic Data Analysis:** This discipline covers various aspects from data storage and organization to automatic analysis algorithms, such as Support Vector Machines, Neural Networks, PCA, etc. It might be classified among others into Data Management, Data Mining (Knowledge Discovery), and Machine Learning.
- **Human Factors:** Human information processing and the human capability of information reception have to be adequately taken into account when analyzing complex data collections. This should be reflected in an appropriate user interface design, a clean requirement analysis and modeling, and perhaps most important an efficient interaction between the human analyst and the computer. Various domains are involved to reach those goals, such as Perception and Cognition Research, Human-Computer Interaction, or (Information) Design.

All of these fields are able to substantially contribute to VA research. In addition, we have to add at least two further aspects to this discussion. First, we have to find a common language to understand the input data and the problems that come with it. Technically, this leads to the more formal definition and standardization of data types, analysis algorithms or visual representations. We must provide **software**

infrastructures for VA that support unification and simple dissemination of data collections as well as the provision of modules/components for visual and/or automatic data analysis. Otherwise, we waste research time and lower software quality as VA modules will be permanently reimplemented by people who need them (Keim et al. 2010, Page 88).

The design of user studies, i. e., performing **evaluations** in order to check the effectiveness of interactive visualizations and thus to validate visualization research, was not in the research focus for a long time. Quite recently, researchers started to evaluate their approaches and ideas with larger subject groups and/or more realistic data sets (Carpendale 2008). One of the reasons for the late consideration of evaluation is that a visualization tool typically consists of a conglomeration of many individual visual representations and interaction techniques. This leads to a challenge as it is very difficult to evaluate a complete software system because of too many variables that cannot be controlled. Evaluating VA systems is even more difficult because of the increased number of disciplines and the diversity of input data sets, potential users, and task to be solved. Thus, new methodologies and, if possible, benchmark data sets have to be developed.

3.1 Data Sources and Types

Data collections that have to be analyzed are often distributed, i. e., they come from diverse data sources and have to be integrated in some way to make use of them. This is still a challenge in data management. In addition, such data may be incomplete, inconsistent and uncertain. A successful VA approach must be able to deal with such problems that are often hard to grasp. It makes also a difference if the data set is existing in a more or less static form, or if it is in the form of a data stream. The latter introduces conceptual and technical problems in analyzing and representing the data because of the permanent adding of new data elements. An example is textual streaming data that is collected from blogs or twitter posts.

A relatively well-understood problem is the (visual) analysis of specific data types, such as multivariate, spatial, time-dependent or network data. Especially in the visualization community, there are many works which provide surveys on how to visually represent specific data types. A nice overview is given in the book chapter (Görg et al. 2007). The authors describe ways to express the most common data types by using visual metaphors and how they can be converted into visual representations suitable for interaction (see Subsection 3.3). For time-oriented data visualization, we refer to the book (Aigner et al. 2011), and for spatio-temporal visual analysis to (Keim et al. 2010, Chapter 5). But note that the visual analysis of a data collection that consists of *many* data types (e. g., quantitative, textual as well as multimedia data) is still not well-understood and additional research is needed to cope with this issue. Another issue is to distinguish between so-called *primary data* (i. e., original input data, directly measured, etc.) and *secondary data* that has been derived or computed from primary data.

3.2 Different Ways of Integration

Several types of an architectural design for a VA system are possible, for instance, the visualization of computational results of the automatic analyses or vice versa the use of visualizations to preselect interesting parts of the data and then to analyze those with automatic techniques. More promising is the use of visualizations to steer the analytical process and to allow the human analyst to interactively choose parts of the data set or to interactively change parameters of the automatic analyses (tightly integrated visualization). Then, results of the analyses can be again visualized, and the analyst can continue with other (visual or automatic) analysis techniques supported by the system. In a similar way, intermediate results of automatic techniques can be visualized too, i. e., automatic analyses can be performed on demand and steered by the analyst. Thus, interactive visualization and exploration is part of the computational machinery itself and supports users in the analysis loop which is crucial for many application fields. Figure 1 shows the described integration of visual and automatic data analysis methods in context of the visual analytics process (extended version of the sense-making loop for Visual Analytics proposed by (Keim et al. 2008)).
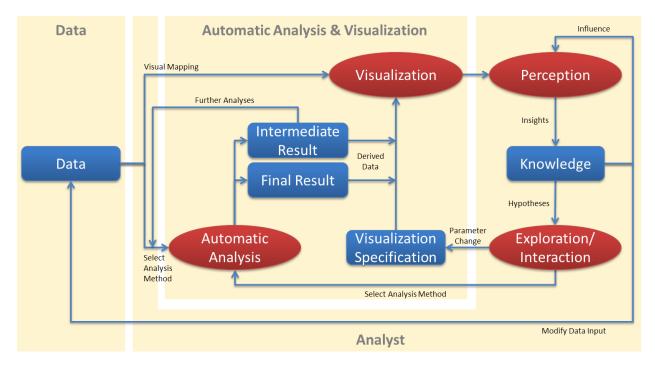
Figure 1: Process diagram to explain the analytical processes in VA. Blue boxes are containers for primary/secondary data, specifications as well as the gained knowledge. Red ovals denote processes. Theoretically, there should be an additional link between Final Result and Knowledge, but we omitted it because of the assumption that the final result should be visualized too. For a detailed discussion on the relationships between data, information and knowledge, the authors refer to Rowley (2007).

In the light of the visual analysis of simulations, the Data is simulation data, and the Automatic Analysis process might be a further simulation run based on modified parameter settings, for example.

3.3 The Role of Interaction

As shown in the previous subsection and Figure 1, the user plays an essential role in the analysis process and the sense-making loop for Visual Analytics. Based on the visualization and the results of automatic analyses, he/she decides to perform further (visual/automatic) analysis steps by interactive exploration of the visualized data. For doing this efficiently, the VA system has to support such interactions that allow the users to decide what they see, how they see it, and what the next analysis steps are. This requirement is also mentioned in the research and development agenda (Thomas and Cook 2005, Page 76).

There are many taxonomies of interaction techniques in the literature which help to better understand the design space of interaction. (Yi et al. 2007) provide a nice overview of the most important taxonomies and propose seven different categories of interaction techniques that connect specific user objectives with those techniques that help accomplishing them. Recently, another work (Heer and Shneiderman 2012) adapted this strategy of identifying interaction categories (based on user objectives) for analytical dialogs. In this paper, we follow their proposal with small modifications as described in the following taxonomy which slightly modifies the one described in (Heer and Shneiderman 2012). Our first high-level category A focuses on the data space and how the data is visually represented. The second category B addresses interacting with the generated visual representations, whereas more general processes of iterative data exploration (e. g., collaboration or modifying interaction histories) are covered by category C.

A.   **Data and View Specification** (data transformations and visual mappings in the InfoVis Reference Model (Card et al. 1999))

    1. *Encode/Visualize:* Users can choose the visual representation of the data records including graphical features, such as color, shape, etc. Data-flow graphs or formal grammars can be used for the specification. Visual representations typically depend on the data types. For instance, hierarchical data sets might be represented by node-link tree layouts or space-filling treemaps.

    2. *Reconfigure:* Some interaction techniques allow the user to map specific attributes to graphical entities. An example is the mapping of attributes in a multivariate data set to different axes in a scatter plot.

    3. *Filter:* This technique is of great importance for visual analysis as it allows the user to interactively reduce the data shown in a view. Popular methods are dynamic queries by using range sliders or picking a set of nodes in a network visualization for further analyses by performing a "lasso" selection.

    4. *Sort:* Ordering of records according to their values is a fundamental operation in the visual analysis process. This is, for example, important in network analysis where nodes might be sorted based on specific centrality values.

    5. *Derive:* Strongly related to the sense-making loop in Subsection 3.2 is the completion of additional computations based on the primary input data. Thus, the user might integrate results of statistical computations (aggregation, medians, . . .) into the data to be visualized.

    6. *Adjust:* Related to the previous interaction type is the modification of parameters for automatic analyses (incl. simulations). In this way, actions in the data space using (configurable) computational methods instead of visual ones can be performed.

B.   **View Manipulation** (view transformations in the InfoVis Reference Model)

    1. *Select:* Selection is often used in advance of a filter operation. The aim is to select an individual object or a set of them in order to highlight, manipulate, or filter out them. Examples include putting a placemark on a virtual map to highlight a spatial area or the specification of attribute ranges in parallel coordinate systems.

    2. *Navigate/Explore:* This important class of interaction techniques typically modify the level-of-detail following the mantra *overview first, zoom and filter, details on demand* (Shneiderman 1996). Well-known approaches are focus&context, overview&detail, zooming&panning, or semantic zooming.

    3. *Coordinate/Connect:* Linking a set of views or windows together to enable the user to discover related items. Brushing and linking techniques (e. g., histogram brushing) are used in almost all VA systems.

    4. *Organize:* Large systems often consist of several windows and workspaces that have to be organized on the screen. Adding and removing views can be confusing to the analyst. Some systems help the user to better overview and to preserve his/her mental map by grouping of views or by assigning specific places where they have to appear.

C.   **Process and Provenance** (facilitating the analytical process)

    1. *Record:* Methods that store and visualize the interaction history performed by the user help to facilitate the iterative analysis process. Undo or redo operations are the most simple examples for such techniques. More advanced ones allow to log complete user sessions (i. e., storage of all user actions such as zooming, filtering, . . .) and can be used to revisit states of the analysis or for collaborative work as colleagues might import a complete analysis trail of someone else.

    2. *Annotate:* Graphical or textual annotations help the analyst to point to elements or regions within the visual representation. These annotations also might be links to other views. Important is that such annotations must be "data-aware" (Heer and Shneiderman 2012), otherwise they may become meaningless in context of interactions.

3. *Share:* Collaboration in VA often occur in practice, but it is still not very well researched. A VA system has to support discussions, dissemination of results, or interactions of several analysts at the same place and the same time (co-located) or at different places and not necessarily at the same time (distributed). Sharing views or publication of visualizations are examples of important requirements for efficient collaboration between many analysts.

4. *Guide:* The specification of workflows is difficult for VA tasks as the related analytical processes are typically non-linear. A result can thus be reached by various interaction and analysis paths within the system. It would be beneficial if a VA system would support "guided analytics to lead analysts through workflows for common tasks" (Heer and Shneiderman 2012).

Of course, it is possible to combine a subset of the aforementioned techniques into a hybrid interaction method (Ward et al. 2010, Page 315). All interaction techniques presented so far can be executed on normal PCs. Specific interaction approaches that support co-located or distributed collaborative visual analysis or that require specific devices, e. g., to provide additional modalities (haptic devices) or more screen space, were not covered. Such topics would widely go beyond the scope of this paper. The interested reader is referred to the literature, such as the book chapter (Fikkert et al. 2007).

4 EXAMPLES FROM BIOLOGICAL SIMULATIONS

Biological systems, a simple cell as well as a complex organism, can be simulated in many ways focusing on different aspects and using different methods. This ranges from the investigation of basic processes, such as the circadian clock or specific signaling pathways, to complex multi-scale modeling of complete organs such as Noble's heart model (Noble 2002). In this section, we focus as an example on the simulation and VA of metabolism.

4.1 Simulation of Metabolism

Metabolism is fundamental to all life processes, for instance, to produce energy and to synthesize substances. Metabolites are transformed into each other by metabolic reactions. Metabolic reactions are usually catalyzed by enzymes and often transporter-mediated (e. g., a substance may be transported from one cellular compartment to another). A large number of reactions occur at any time in organisms, and the product of one reaction is usually used by another reaction. Thus, metabolic reactions are strongly interlinked and build metabolic networks. More formally, a metabolic network is a hypergraph: the vertices of the network represent the metabolites, the hyperedges represent the reactions. For simulation purposes, the metabolic network is often represented as bipartite graph where additionally to the vertices representing metabolites, the reactions are vertices and edges are binary relations connecting the metabolites of a reaction with the related reaction vertex.

    Simulating metabolism is an important method to understand or even to help manipulating life processes. It has been successfully applied to many studies of bacteria, yeast, plants and other organisms. The behavior of a metabolic system can be simulated with different modeling methods such as household models, stoichiometric models, and kinetic models; all methods which describe different approaches to quantify and simulate fluxes in metabolic networks. The widely used stoichiometric modeling approach is based on the description of the structure of a metabolic network including the stoichiometries of enzymatic reactions, that is the relative quantities of metabolites in a reaction. Stoichiometric models can be used, for example, to find different routes from one metabolite to another, to calculate flux distributions for optimal yield, to investigate the effect of genetic or environmental changes onto metabolic fluxes, and to predict the effect of additional reactions in the network. Two typical simulation approaches based on stoichiometric models are Flux Balance Analysis and Petri nets:

- **Flux Balance Analysis:** Flux Balance Analysis (FBA) (Varma and Palsson 1994) can be performed for a stoichiometric model with additional information such as uptake rates of metabolites and the
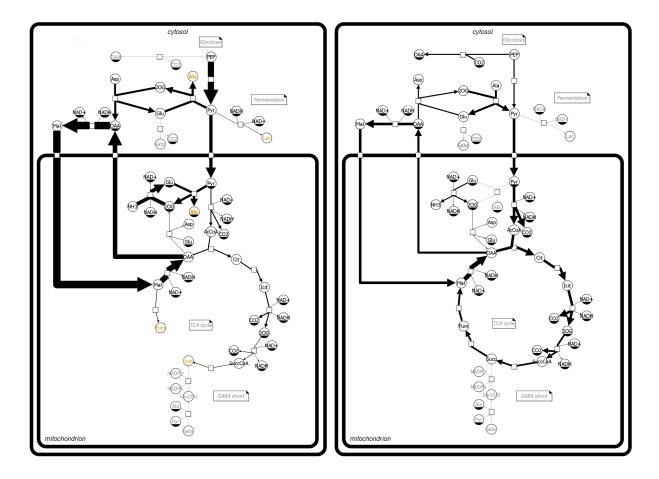
Figure 2: Visualization example of metabolic fluxes computed with FBA. Flux distribution is shown under two different conditions (left and right part of the figure). The circles represent metabolites, the small rectangles represent reactions, large rounded rectangles show the different cellular spaces (compartments), and the edge width represents the metabolic flux over this edge.
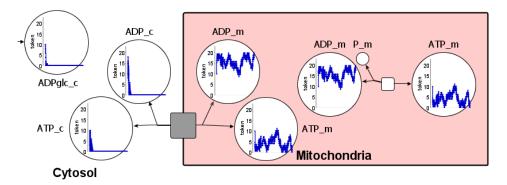

Figure 3: Example of a visualization of metabolite concentrations (which corresponds to the related fluxes) using a Petri net-based simulation. As in Figure 2, circles represent metabolites, and rectangles represent reactions. Here, the change of metabolite concentrations over time is shown as plots.

composition of the produced biomass. It is a constraint-based modeling approach which predicts metabolic steady-state fluxes via applying mass balance and other constraints to the stoichiometric model. The predicted flux through each reaction is optimal for a certain criterion such as the maximization of growth. An example of the results of FBA is visualized in Figure 2. Flux Balance Analysis is a widely used method that has been applied to stoichiometric models of many organisms.

- **Petri Nets:** Petri nets, which have been defined by (Petri 1962), are a mathematical formalism for the representation and analysis of causal systems within concurrent processes. They can also be used for the simulation of stoichiometric models (cf. Figure 3) and have been first applied to metabolic systems around 20 years ago (Hofestädt 1994; Reddy et al. 1993). A review of the use of Petri net simulation for molecular biology, especially metabolism, can be found in (Koch et al. 2011).

The automatic visualization and interactive exploration of the structure of metabolic networks is an active research area since several years. To our knowledge, the first layout algorithm for metabolic networks has been described nearly 20 years ago (Karp and Paley 1994), and several more were developed since then. The question of a common language to understand the input data has been addressed in the last years, and the Systems Biology Graphical Notation (SBGN) (Le Novère et al. 2009) has been developed as a standard for the graphical representation of biological networks and cellular processes including metabolism. Also interactive exploration methods for metabolic networks have been presented, for example, in (Rohrschneider et al. 2010; Rohrschneider et al. 2010; Streit et al. 2008; Klukas and Schreiber 2007; Kono et al. 2009; Jusufi et al. 2012). There are also several tools to create and simulate metabolic systems using FBA or Petri net approaches (a recent review which also covers related aspects in metabolic engineering is (Copeland et al. 2012)) as well as tools to visualize precomputed fluxes in metabolic systems (for a comparison of current systems the interested reader is referred to (Rohn et al. 2012)). However, the visual investigation of metabolic fluxes is a relatively new area and only few of the above mentioned systems support both, the integrated simulation using FBA or Petri nets and the interactive visualization of metabolic models and simulation results.

4.2 Visual Analytics During Model Reconstruction

Model reconstruction is the building of a stoichiometric model out of information from literature, databases, experiments, and personal knowledge. It aims on producing a balanced (i. e., stoichiometrically correct and dead end free) model which represents the current knowledge about metabolism of a specific organism, organ, tissue or cell. Often, not the complete metabolism (also called *genome scale model*) but only specific subparts are reconstructed.

The process of model reconstruction heavily involves human interaction. Either the complete model is manually build, or it is manually refined from existing or computer-generated models. The model represents the Data container in Figure 1 (text set in sans-serifs refers in the following to elements in this figure). Model reconstruction is an iterative process where subsequent models are tested (based on Automatic Analysis) and refined (Modify Data Input). Automatic Analysis includes finding unwanted metabolic sinks, gaps in pathways, unbalanced reactions, and missing transporters [A.5] (interaction categories according to Subsection 3.3 are given in squared brackets). These steps lead to Intermediate Results and the Visualization of the current model including highlighting problems or errors [A.1]. Although models are exchanged in computer-readable formats (e. g., SBML), a human-readable representation of the model as network diagram (Visualization) is important to support not only a better understanding, but also to investigate the structure of the model, or to find gaps (missing reactions). As a simple example, an Automatic Analysis may find all unconnected metabolites and the subsequent Visualization may highlight them, thereby indicating to a user that these parts of the model need further investigation. Typical interaction techniques (Exploration/Interaction which leads to a modified Visualization Specification) used in the process of model reconstruction are zooming&panning [B.2] (e. g., genome scale models may contain

thousands of metabolites and reactions), filtering of subsystems [A.3] (e. g., compartments or specific pathways), focus&context [B.2], and visualization of analysis results [A.1, A.2, A.5] (e. g., reachability, balance checking, and unused metabolites). The whole cycle of Automatic Analysis, Visualization and Exploration/Interaction including the discussed steps will be repeated until a valid model has been obtained. It should be noted that in later states of this iterative process of model reconstruction this goes hand in hand with simulation. Simulation results are initially often not corresponding to knowledge and lead to model improvement, see the next Subsection 4.3.

In conclusion, analysts interactively build stoichiometric models, use visualization to understand and evaluate the models and their current problems (here, we mean the evaluation of the metabolic model, not the evaluation to check the effectiveness of interactive visualizations) during the iterative model building process, and thereby employ Visual Analytics techniques in the course of the model reconstruction process.

4.3 Visual Analytics for Simulation Result Analysis

Model simulation leads to simulation results which have to be analyzed to derive knowledge or build novel hypotheses about the metabolic system. It aims on understanding the metabolic system, i. e., what fluxes occur, which branches are used, what bottlenecks are there, how do fluxes change under different input parameters, how do fluxes change under structural changes of the model, etc. The analytical process follows the sense-making process explained in Figure 1. This interactive process of simulation result investigation finally leads to an evaluation of the model, better understanding of the metabolism, and the possibility to predict changes of the metabolism by manipulating and simulating the model.

Here the Data is the stoichiometric model and additional data, such as in-fluxes into the system and biomass composition. Automatic Analysis is the simulation of the input model under different scenarios, for example, finding an optimal flux distribution, computing flux variability [A.5] or investigation the effect of restrictions such as temporarily removing reactions [A.6]. The Intermediate Results are fluxes over reactions or metabolite concentrations, which are commonly shown in the context of the model (Visualization). Human interaction is strongly involved in the process of analyzing simulation results as the investigation of model scenarios is an interactive process (e. g., how do fluxes change if parameters of the model are changed). Although simulation results can also be analyzed statistically, a human-readable representation of the simulation is a prerequisite for intuitive flux data exploration in the context of metabolic networks. This applies not only to stoichiometric modeling and FBA or Petri net simulation as in Figures 2 and 3, but also to other simulation scenarios. Typical interaction techniques (Exploration/Interaction that results in changed parameters for the visualization (Visualization Specification), but also for the simulation) are again zooming&panning [B.2], filtering [A.3] (e. g., of subsystems, of reactions with specific flux ranges), and focus&context [B.2]. For example, a user may initially compute a flux distribution [A.5], investigates this in different parts of the model [A.1] using Exploration/Interaction and decides to analyze the effect of temporarily removing a reaction [A.6] (Automatic Analysis). This results in a new flux distribution which may lead to a better representation (Visualization) of observed effects in metabolism (new Knowledge). As an result, the user may update the model accordingly to reflect this new knowledge (Modify Data Input) and based on this starts the analysis cycle again.

To sum up, analysts run model simulations, investigate simulation results interactively, and use visualization to understand the metabolism represented by the model (investigate changing parameters, *in silico* knockouts, etc.).

5 CONCLUSIONS

In this paper, we provided the simulation community an overview of the most important ideas and concepts of Visual Analytics. A brief presentation of the most important books and events that cover this field gives the reader an entry point to further literature and general information or interesting initiatives. We illuminated the diverse facets of Visual Analytics from related research fields to the role of interaction

in the analysis process. After this more theoretical and conceptual discussion on this emerging field, we continued with concrete examples from biological simulations (especially of metabolism) and motivated the importance of the human in the analysis loop in order to achieve more knowledge and understanding of the simulation results and underlying models.

We hope that we could convince the reader that interactivity plays an important role in the analysis process, and that Visual Analytics systems have to support the human analyst by a smart combination of interactive visualizations and automatic analysis methods.

REFERENCES

Aigner, W., S. Miksch, H. Schumann, and C. Tominski. 2011. *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. Springer.

Bryant, R. E., R. H. Katz, and E. D. Lazowska. 2008. "Big-data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society". In *Computing Research Initiatives for the 21st Century*. Computing Research Association.

Card, S., J. Mackinlay, and B. Shneiderman. (Eds.) 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

Carpendale, S. 2008. "Evaluating Information Visualizations". See Kerren, Stasko, Fekete, and North (2008), 19–45.

Ceglar, A., J. F. Roddick, and P. Calder. 2003. "Managing Data Mining Technologies in Organizations". Chapter Guiding Knowledge Discovery through Interactive Data Mining, 45–87. Hershey, PA, USA: IGI Publishing.

Copeland, W. B., B. A. Bartley, D. Chandran, M. Galdzicki, K. H. Kim, S. C. Sleight, C. D. Maranas, and H. M. Sauro. 2012. "Computational Tools for Metabolic Engineering". *Metabolic Engineering* 14:270–280.

Fikkert, W., M. D'Ambros, T. Bierz, and T. Jankun-Kelly. 2007. "Interacting with Visualizations". See Kerren, Ebert, and Meyer (2007), 77–162.

Gehlenborg, N., S. I. O'Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, D. Tenenbaum, and A.-C. Gavin. 2010. "Visualization of Omics Data for Systems Biology". *Nature Methods* 7 (3 Suppl): S56–S68.

Görg, C., M. Pohl, E. Qeli, and K. Xu. 2007. "Visual Representations". See Kerren, Ebert, and Meyer (2007), 163–230.

Heer, J., and B. Shneiderman. 2012, April. "Interactive Dynamics for Visual Analysis". *Commun. ACM* 55 (4): 45–54.

Hofestädt, R. 1994. "A Petri Net Application to Model Metabolic Processes". *Syst. Anal. Mod. Simul.* 16 (2): 113–122.

Jusufi, I., C. Klukas, A. Kerren, and F. Schreiber. 2012. "Guiding the Interactive Exploration of Metabolic Pathway Interconnections". *Information Visualization* 11 (2): 136–150.

Karp, P. D., and S. M. Paley. 1994. "Automated Drawing of Metabolic Pathways". In *Proc. International Conference on Bioinformatics and Genome Research*, edited by H. Lim, C. Cantor, and R. Bobbins, 225–238.

Keim, D., G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. 2008. "Visual Analytics: Definition, Process, and Challenges". See Kerren, Stasko, Fekete, and North (2008), 154–175.

Keim, D., J. Kohlhammer, G. Ellis, and F. Mansmann. (Eds.) 2010. *Mastering The Information Age – Solving Problems with Visual Analytics*. Eurographics Digital Library.

Kerren, A., A. Ebert, and J. Meyer. (Eds.) 2007. *Human-Centered Visualization Environments*. LNCS Tutorial 4417. Heidelberg: Springer.

Kerren, A., J. T. Stasko, J.-D. Fekete, and C. North. (Eds.) 2008. *Information Visualization: Human-Centered Issues and Perspectives*, Volume 4950 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer.

Klukas, C., and F. Schreiber. 2007. "Dynamic Exploration and Editing of KEGG Pathway Diagrams". *Bioinformatics* 23 (3): 344–350.

Koch, I., W. Reisig, and F. Schreiber. 2011. *Modeling in Systems Biology: The Petri Net Approach*. Springer.

Kono, N., K. Arakawa, R. Ogawa, N. Kido, K. Oshita, K. Ikegami, S. Tamaki, and M. Tomita. 2009. "Pathway Projector: Web-based Zoomable Pathway Browser Using KEGG Atlas and Google Maps API". *PLOS One* 4 (11): e7710.

Le Novère, N., M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M. Aladjem, S. M. Wimalaratne, F. T. Bergman, R. Gauges, P. Ghazal, K. Hideya, L. Li, Y. Matsuoka, A. Villéger, S. E. Boyd, L. Calzone, M. Courtot, U. Dogrusoz, T. Freeman, A. Funahashi, S. Ghosh, A. Jouraku, S. Kim, F. Kolpakov, A. Luna, S. Sahle, E. Schmidt, S. Watterson, G. Wu, I. Goryanin, D. B. Kell, C. Sander, H. Sauro, J. L. Snoep, K. Kohn, and H. Kitano. 2009. "The Systems Biology Graphical Notation". *Nature Biotechnology* 27:735–741.

Noble, D. 2002. "Modeling the Heart: From Genes to Cells to the Whole Organ". *Science* 295:1678–1682.

Petri, C. A. 1962. *Communication with Automata (in German)*. Institut für Instrumentelle Mathematik, Bonn: Schriften des IIM 3.

Reddy, V. N., M. L. Mavrovouniotis, and M. N. Liebman. 1993. "Petri Net Representations of Metabolic Pathways". In *Proc. International Conference on Intelligent Systems for Molecular Biology (ISMB'93)*, edited by L. Hunter, D. Searls, and J. Shavlik, 328–336.

Rohn, H., A. Hartmann, A. Junker, B. H. Junker, and F. Schreiber. 2012. "FluxMap: A VANTED Add-on for the Visual Exploration of Flux Distributions in Biological Networks". *BMC Systems Biology* 6:33.

Rohrschneider, M., C. Heine, A. Reichenbach, A. Kerren, and G. Scheuermann. 2010. "A Novel Grid-based Visualization Approach for Metabolic Networks with Advanced Focus&Context View". In *Proc. International Symposium on Graph Drawing (GD '09)*, edited by D. Eppstein and E. R. Gansner, Volume 5849 of *LNCS*, 268–279: Springer.

Rohrschneider, M., A. Ullrich, A. Kerren, P. F. Stadler, and G. Scheuermann. 2010. "Visual Network Analysis of Dynamic Metabolic Pathways". In *Proceedings of the 6th International Conference on Advances in Visual Computing – Volume Part I*, edited by G. Bebis, R. Boyle, B. Parvin, D. Koracin, and R. Chung, ISVC'10, 316–327. Berlin, Heidelberg: Springer.

Rowley, J. 2007. "The Wisdom Hierarchy: Representations of the DIKW Hierarchy". *Journal of Information Science* 33 (2): 163–180.

Shneiderman, B. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". In *Proceedings of the IEEE Symposium on Visual Languages (VL '96)*, 336–343: IEEE CS Press.

Short, J. E., R. E. Bohn, and C. Baru. 2011. "How Much Information? 2010 Report on Enterprise Server Information". Technical report, Global Information Industry Center, UC San Diego, USA.

Streit, M., M. Kalkusch, K. Kashofer, and D. Schmalstieg. 2008. "Navigation and Exploration of Interconnected Pathways". *Comput. Graph. Forum* 27 (3): 951–958.

Thomas, J. J., and K. A. Cook. 2005, August. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE CS Press.

Thomas, J. J., and K. A. Cook. 2006. "A Visual Analytics Agenda". *Computer Graphics and Applications, IEEE* 26 (1): 10–13.

VA Community 2012. "Visual Analytics Community – Discovery Through Collaboration and Sharing". Accessed October 22, 2012. http://www.vacommunity.org.

Varma, A., and B. O. Palsson. 1994. "Review: Metabolic Flux Balancing: Basic Concepts, Scientific and Practical". *Bio/Technology* 12:994–998.

VisMaster 2012. "European Coordination Action Project on Visual Analytics". Accessed October 22, 2012. http://www.vismaster.eu/.

VisMaster Video 2012. "Visual Analytics – Mastering the Information Age". Accessed October 22, 2012. http://www.youtube.com/watch?feature=player_embedded&v=5i3xbitEVfs.

Visual-Analytics.EU 2012. "European Web Portal on Visual Analytics". Accessed October 22, 2012. http://www.visual-analytics.eu/.

Ward, M., G. Grinstein, and D. A. Keim. 2010. *Interactive Data Visualization: Foundations, Techniques, and Application*. A.K. Peters, Ltd.

Wong, P. C., and J. Thomas. 2004, sept.-oct.. "Visual Analytics". *Computer Graphics and Applications, IEEE* 24 (5): 20–21.

Yi, J. S., Y. a. Kang, J. Stasko, and J. Jacko. 2007, November. "Toward a Deeper Understanding of the Role of Interaction in Information Visualization". *IEEE Transactions on Visualization and Computer Graphics* 13 (6): 1224–1231.

AUTHOR BIOGRAPHIES

ANDREAS KERREN received the B.S. and M.S. degrees as well as his PhD degree in Computer Science from the Saarland University, Saarbrücken (Germany). In 2008, he achieved his habilitation (docent competence) from Växjö University (Sweden). Dr. Kerren is currently a Full Professor in Computer Science at the School of Computer Science, Physics and Mathematics, Linnaeus University (Sweden), where he is heading the research group for Information and Software Visualization, called ISOVIS. His main research interests include the areas of Information Visualization, Software Visualization, Visual Analytics, and Human-Computer Interaction. Dr. Kerren is a member of the ACM, the Eurographics Association, the IEEE Computer Society and of various networks such as the IFIP Working Group 13.7: HCI and Visualization. His email address is andreas.kerren@lnu.se.

FALK SCHREIBER graduated, obtained a PhD and a habilitation in Computer Science from the University of Passau (Germany). He worked as a Research Fellow and Lecturer at the University of Sydney (Australia) and has been head of a research group at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben (Germany). In 2007, he has been appointed as Professor of Bioinformatics at the Martin Luther University Halle-Wittenberg (Germany) and as Bioinformatics Coordinator at the IPK Gatersleben. His research focuses on the modeling, analysis and visualization of biological networks and related data, the simulation of metabolic processes, and Visual Analytics of multimodal and multidimensional biological data. His email address is schreibe@ipk-gatersleben.de.