

Article type: Overview

# An Overview on Interactive Visual Data Mining Techniques for Knowledge Discovery

Frederic Stahl

University of Reading, School of Systems Engineering, F.T.Stahl@reading.ac.uk

Bogdan Gabrys

Bournemouth University, The School of Design, Engineering & Computing

Mohamed Medhat Gaber

University of Portsmouth, School of Computing

Monika Berendsen

Evonik Industries AG

## Abstract

In the past decade the analysis of data has faced the challenge of dealing with very large and complex datasets and the real time generation of data. Technologies to store and access these complex and large datasets are in place. However, robust and scalable analysis technologies are needed in order to extract meaningful information from these datasets. The research field of *Information Visualisation* and *Visual Data Analytics* addresses this need. Information visualisation and data mining are often used complementary to each other. Their common goal is the extraction of meaningful information from complex and possibly large data. However, whereas data mining focuses on the usage of silicon hardware, visualisation techniques also aim to access the powerful image processing capabilities of the human brain. This article highlights the research on data visualisation and Visual Analytics techniques. Furthermore we highlight existing Visual Analytics techniques, systems and applications including a perspective on the field from the chemical process industry.

## Keywords

Visual Analytics, Data Mining, Information Visualisation, Data Stream Mining

## Introduction

It is predicted that the digital universe in the year 2020 will be 44 times as large as it was in the year 2009, where it was estimated to be 800,000 petabytes in size (1). Computer

systems record our daily lives (2), for example CCTV cameras and GPS enabled navigation devices such as smart phones record our whereabouts; loyalty card and credit card issuers record our consumer behaviour; and social media applications such as facebook or email capture our social life (2). However, also in the scientific arena large quantities of data are stored and need to be analysed. For example in astronomy, data bases comprise terabytes of image data (see the GSC-II (3) or the Sloan Digital Sky survey (4)); or in bioinformatics the human genome project stores the genetic blueprint of the entire human body comprising 20,000 - 25,000 genes and 3 billion chemical base pairs (5). Hardware technology allows us to not only capture but also store this data. Hence the commercial demand in developing and improving analysis techniques to extract meaningful information from the digital universe is growing (6). This leads to the development of novel and scalable Visual Analytics techniques.

According to Xie (7), in the recent years data mining researchers have started to regard visualisation techniques as a critical aspect of the decision making and data analysis process. Visualisation can help analysts to visually discover different kinds of patterns such as clusters, relationships and associations. The visual analysis capabilities of the human brain can be combined with the silicon computer hardware in order to improve the scalability and accuracy of the discovery of patterns from large and complex data. Keim and colleagues lay out that this combined usage of silicon and biological hardware should be augmented with the human expert's knowledge in an interactive way (8). This is described as Visual Data Analytics or short *Visual Analytics*.

The aim of this overview article is to discuss the role of Visual Analytics for data analysis in the overall Knowledge Discovery from Data (KDD) process. Furthermore this article highlights existing Visual Analytics software systems and applications. One particularly relevant application of data visualisation and Visual Analytics techniques is the chemical process industry. This industry offers applications in the research for new products and production methods, as well as for monitoring the current production status and trends on large chemical plants. One of the biggest global players in the chemical industry is Evonik Industries AG and this article provides an overview on the use of Visual Analysis techniques in this company.

The article is organised as follows: First the article discusses *Visual Data Analytics* and how it fits into the overall Knowledge Discovery from Data process. Then typical components and concepts often used in such Visual Data Analytics software systems are discussed. Next some actual Visual Data Analytics software systems are highlighted followed by a review of visualisation techniques used in data stream mining. Then the article describes Visual Data Analytics applications in the chemical process industry amongst other applications. The article is concluded by highlighting major challenges and providing an outlook into the future of Visual Data Analytics.

## The Interactive Process of Visual Data Mining and Knowledge Discovery from Data

This section highlights the basic idea of the combination of visualisation and data analytics techniques and discusses the basic principles involved.

Information Visualisation is defined as *'the use of computer-supported, interactive, visual representations of abstract data to amplify cognition'* (9). Information Visualisation aims to gain understanding of data using graphical representations in order to access the powerful image processing capabilities of the human brain. This is described in the information seeking mantra as *'overview first, zoom/filter, details on demand'* (10).

Visual Analytics has been defined as the *'science of analytical reasoning facilitated by interactive visual interfaces'* (11). Visual Analytics is based on the same ideas as Information Visualisation. However, Visual Analytics also comprises the incorporation of automatic analysis methods prior, after and during the usage of interactive visual representations (8).

### The Principle of Visual Analytics

According to (66) Visual Analytics as a field of research started in 2005 with (11) as commencement. Keim describes the Visual Analytics processes as a feedback loop comprising different stages and transitions (47) as illustrated in Figure 1. The first stage is the gathering of data. The data may come from different data sources and hence needs to be integrated before further analysis methods can be applied. This step also comprises pre-processing of data such as data normalisation and data cleaning. This pre-processing and integration is represented by the transformation arrow.

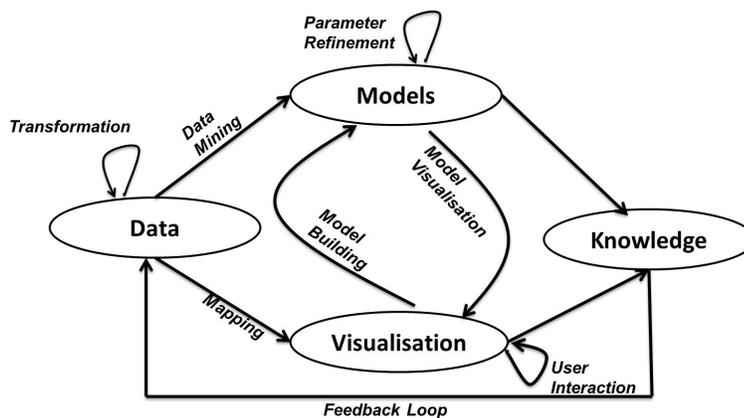


Figure 1: The process of Visual Analytics according to (47).

After this transformation the analyst applies either visualisation techniques or automated analysis methods. Automated analysis methods can comprise statistical and data mining techniques. Visualisations are then used to evaluate the findings but also to refine the automated analysis methods. This alternation between automatic analysis and information visualisation is the main characteristic of Visual Analytics and the difference when compared to just Information Visualisation. This is also reflected in the mantra of Visual Analytics ‘*Analyse First - Show the Important - Zoom, Filter and Analyse Further - Details on Demand*’ (8). It is important to optimise the visualisations for the human visual system in order to create an effective Visual Analytics system; also user interaction methods such as focus and context are important for scientific visualisation (53).

## Visual Analytics and KDD

However, data mining for model building is just one step of the overall Knowledge Discovery from Data (KDD) process, and Figure 2 describes the KDD process augmented with the Visual Analytics process as discussed in the ‘*The Principle of Visual Analytics*’ section.

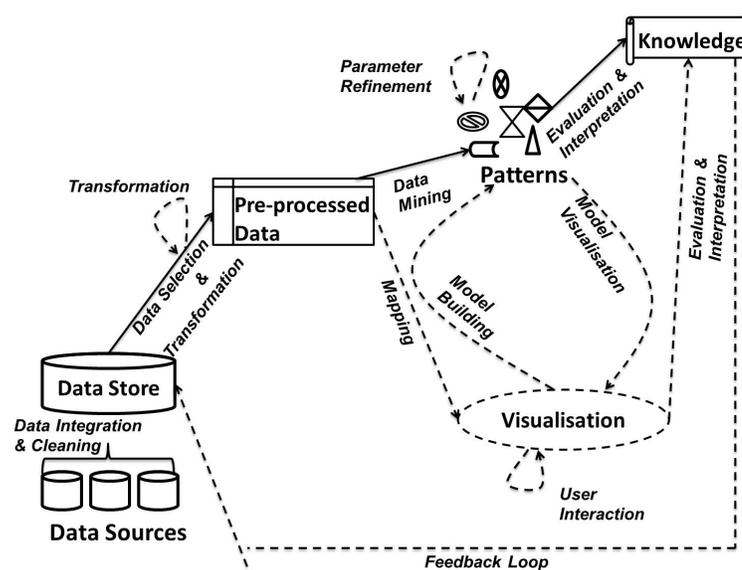


Figure 2: The Knowledge Discovery from Data process augmented with the Visual Analytics process. The dashed components and lines are the augmented elements from Visual Analytics.

The dashed printed elements are taken from the Visual Analytics process, whereas the solid printed elements correspond to the KDD process. Please note that none of the

KDD steps have been removed. Also note that as opposed to the KDD process, the Visual Analytics process is interactive through interaction with the user of the system and the feedback loop is feeding changes back to the data input (48) as shown in Figure 1.

The following steps are extracted from Figure 2 and describe the main activities involved in the KDD process augmented by the process of Visual Analytics as it is depicted in Figure 1:

- *Data Integration and Cleaning*: comprises the integration of different data sources into one data store or warehouse.
- *Data Selection and Transformation*: prior to the data mining or analysis task the data needs to be selected from the data store. When selecting data, this data may be transformed, for example normalised, cleaned, reduced or converted into a certain format for the data mining step or for the visualisation. In the original KDD process the data analyst is responsible for this. However, after incorporating of Visual Analytics this step may be automated with user interaction information and the feedback used instead. Two versions of the data may be generated, one in an appropriate format for the data mining step and one for visualisation purposes.
- *Data Mining*: intelligent methods are applied to the pre-processed data in order to extract useful patterns, such as classification or association rules. However these models or patterns are refined by the user through interaction with their visual representation.
- *Evaluation and Interpretation*: in the original KDD process, models derived in the data mining step need to be examined in order to identify the interesting aspects of the extracted patterns and to find a meaningful representation. However, the application of the visualisation methods can provide direct insights to the analyst.

Information Visualisation and Visual Analytics are similar approaches as they both aim to combine the human brain and silicon hardware in order to analyse data. However, having made the distinction between both terms and putting Visual Analytics in the context of KDD, one must note that these terms are often used interchangeably in the literature with Information Visualisation actually referring to Visual Analytics or Visual Data Analytics. This paper is no exception. A comprehensive book on Information Visualisation is (26) and on Visual Analytics is (47).

Another confusion often made in the literature is the interchangeable use of the terms Visual Online Analytical Processing (Visual OLAP) and Visual Analytics. OLAP refers to the ad hoc exploration of multidimensional data volumes where each feature represents a dimension. One common task in OLAP is to find intersections and relationships between dimensions, analyse and visualise them. Visual OLAP additionally uses visualisations in order to interact with the analytical algorithms and to refine

them. This is very similar to Visual Analytics. However, Visual Analytics is not limited to multidimensional data.

The next section highlights some existing data mining software systems that incorporate the ideas of Visual Analytics and Information Visualisation. Some of these systems are of commercial nature and some are free.

## **Typical Concepts and Components of Visual Data Analytics Systems**

This section discusses typical components and concepts that are often found in software systems that are more focussed on the visual aspects of knowledge discovery from data. The visual concepts outlined in this section can be found in many Visual Data Analysis software systems. As it has been highlighted in the previous section, Visual Data Analysis or Visual Analytics is more than just static data representations. The user rather manipulates the data and data representations through interactions with multiple and diverse views. With the expression *view* we refer to a graphical data *representation*. We may use the word view and representation interchangeably in this article. Some visual and interaction concepts that can be found in many visual data analysis systems are highlighted below. When we talk about views, we talk about a visual component in the graphical user interface that visualises some aspects of the system, such as data, configurations, data mining models, data mining results, etc. In the case when the view displays data we also call it data representation.

### **Typical User Interface Concepts**

Visualisations can be as versatile as the number of possible applications or even datasets used. In fact many data mining and visual data analysis systems such as (19; 67) are open source or at least allow to integrate new views in the form of plug-ins, and thus allow users to integrate their own individual application tailored views. Hence the section titled '*Typical Concepts and Components of Visual Data Analytics Systems*' highlights visual concepts and views in a more general sense.

Visual Analytics often comprises the usage of multiple views, which requires a well designed and intuitive user interface, taking into consideration the display and arrangement of the visualisation and allow the user to interactively parameterise views (67). Most visual data analysis tools are based on the same or a similar layout compared with the 'Eclipse Rich Client Platform' (18). This layout or similar ones are used in many data mining or data visualisation tools for example in (19; 67; 62). This has the advantage that the visual layout of the framework and basic interactions are already familiar with many potential users. However, some tools use their individual visualisation techniques different to those used in Eclipse, for example Weka's explorer (2). Figure 3 shows a typical Eclipse based GUI comprising the basic views labelled in black letters in the screen shot, however, additional views are possible. The *Project*

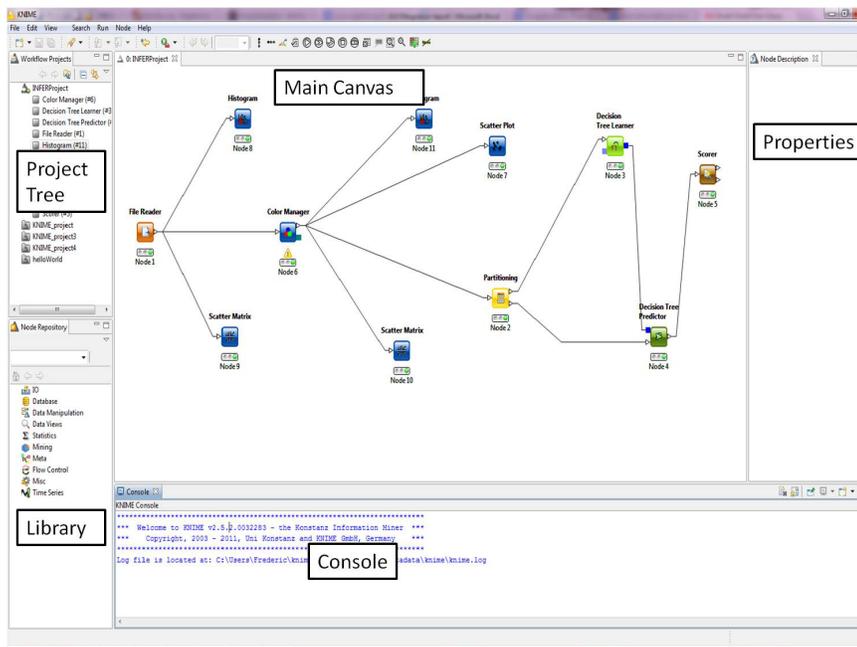


Figure 3: A typical Eclipse RPC Graphical User Interface taken from the KNIME (19) data mining software.

*Tree* is usually a hierarchical navigator through the project; the *Library* offers to select tools, views, data etc. for data mining and visualisation projects; the *Properties* view usually displays information about selections in the project tree or in the main canvas; the main canvas shows a visual representation of the project, usually in the form of a workflow. However, other visualisations are possible in the main canvas, such as the graph representation in the CGV tool highlighted in the ‘*Available Visual Data Analysis Software Systems*’ section. The *Console* view produces textual and tabular output such as error messages or warnings.

## Interaction Techniques

According to the authors of (29) interaction in the context of *Human Computer Interaction* can be described as the communication between the user and the system. In the context of Figure 1 user interaction is essential for the visual exploration of data and the adjustment of data mining models. Only well designed interaction with the view and the data mining models can help the user to browse and select subsets of the data; adjust the visual mapping of the data; visually modify the parameter setting and adjust the visual mapping of the data mining model and the associated results.

In this sense the user performs a *dynamic interaction*, a two way process of feeding information into the data mining system but also retrieving new or modified informa-

tion through the visualisation. This is in contrast to a *passive interaction* using static images that can be observed, rotated, zoomed and un-zoomed in order to enhance the user's mental model on the data (65). According to the authors of (13) an interaction technique is defined as a way of using a physical input/output device in order to perform a task on using a human-computer dialogue. Some basic interaction concepts can be found in many visual data analysis systems and are discussed here in general terms rather than looking into application specific techniques. For more information about data/information visualisation we refer the interested reader to (72).

The authors of (74) conducted an exhaustive review of existing software systems and the literature about information visualisation and hence data visualisation. In particular, 59 papers, 51 systems and 311 identified individual interaction techniques have been reviewed (74). However, the authors further point out that different interaction techniques aim to achieve the same or similar goals and hence they established a categorisation of these techniques based on the notion of user intent in using these techniques. These categories are described below, which are (a) Select, (b) Explore, (c) Reconfigure, (d) Encode, (e) Abstract/Elaborate, (f) Filter and (g) Connect. For visual interaction examples we refer to the articles cited within the description of the interaction techniques.

*a) Select:* This technique of interaction enables the user to mark interesting data items in a view. This allows the user to easily identify and keep track of the data items of interest if the representation of the data is changed, especially if there are many items represented in the view. An example of select that can be found in Weka and KNIME is the scatter matrix (2; 19). A scatter matrix is a matrix of plots of each feature against each other. Selected data points in one plot can be highlighted across all plots in the scatter matrix. *Select* can be combined with other techniques, i.e. *Reconfiguration* in order to see where data items move when rearranging representations. A well known example of *Select* used together with *Reconfiguration* is the 'placemark' feature of Google Earth (44). 'Placemark' allows to select a geographical location, rotate and zoom the view and easily return to the marked location.

*b) Explore:* This technique enables the user to view and inspect a different subset of the data. This is useful especially if the data is very large and hence screen sizes as well as the processing capabilities of the human brain limit the data representation and processing as a whole. The user can explore a subset, gain insights and move on to explore further data. Yi's review further highlight two commonly used approaches to *Explore* interaction, namely *Panning* and *Direct-Walk*. In photography panning refers to the movement of a scene in front of a fixed camera. In the context of information visualisation or visual analytics the scene is the data and the camera the user's eye. The simplest way of moving the scene could be via scroll bars (74). *Direct-Walk* allows the user to move the focus of the view from one position in the data structure to another, i.e. using hyperlinks such as in web browsers.

*c) Reconfigure:* This technique enables the user to change the data representation by changing its spatial arrangement. Thus the user is presented with different views of the data which helps to uncover hidden relationships. For example in scatter plots some of the data elements may have similar or the same visualised numerical values, which

results in the data elements to overlap in the plot. In this case jitter can be used to make hidden data elements visible. Jitter refers to shifting the data elements randomly in the display space by a small amount in order to avoid the overlap of data elements (55). For example the jitter technique is used by the authors of (55) and in the Spotfire (64) visual analytics software. But also the rotation of 3D scatter plots may reveal overlapping data elements.

*d) Encode:* This technique enables users to alter the representation of data elements. For example altering their size, colour or shape. Altering the representation is intended to increase the human cognition in terms of understanding relationships in the data and the distribution of data elements. In Yi's review two widely used encoding approaches are highlighted, *changing the type of the data representation* and *interaction to change the encoding of the data elements* such as colour, size or shape (74). Changing the type of the data representation is intended to reveal new aspects of relationships between data elements. A Tool that allows to change the data representation is i.e. (64). Changing the encoding of data elements through user interaction is intended for the user to find colour schemes most suitable to discover distributions of multiple variables or features. Tools that allow the user to change the encoding system based on colour or similar encoding mechanisms are for example (37; 73).

*e) Abstract/Elaborate:* This technique enables the user to adjust the level of detail of the data representation. This allows the user to view the data in a wider context or in a more detailed view on demand often with many different levels of detail in between the context and the individual data elements. This can be realised in many different ways, i.e. tool tip texts that appear when the user moves the mouse cursor over a particular data element. However, tool tip texts usually only allow two levels of detail. More sophisticated approaches are needed i.e. graph or tree views. The expansion and collapsing of graph or tree nodes to reveal and hide sub trees or sub graphs is for example implemented in the CGV system (67).

*f) Filter:* This interaction technique enables the user to define conditions that change the set of data items displayed in the current view. The data items that are 'filtered out' remain unchanged but are hidden or displayed in a different way. Usually the filter interaction is complemented with a reset facility that allows to recover the hidden or differently displayed data items. A popular approach of performing filter operations in visualisation tools is the usage of dynamic query controls to select conditions/ranges such as check boxes (20; 74). More specialised controls exist, for example in the *TimeSearcher* tool (41) that allows the data analyst to define conditions graphically using 'timeboxes' and 'angular queries'. 'Timeboxes' are used for filtering multiple univariate time-series profiles according to time and value ranges; and 'angular queries' are used for filtering according to the time series value rate of change in a given time frame (41).

*g) Connect:* This interaction technique enables the user to highlight relationships between data items. Highlighting can happen within the same view or across different data representations. An example of highlighting between different views is KNIME's (19) *line plot* and *scatter plot*, data items highlighted in either of these plots are cross highlighted in the other active plot(s) as well. An example for highlighting within the

same view is i.e. CGV's magic eye view (67). The magic eye projects a graph's hierarchy in the form of a tree onto a 3D hemisphere. CGV extends the magic eye view in order to visualise cross edges among selected hierarchy nodes by spanning an arch around the 3D hemisphere.

At this stage we would like to highlight that not all interaction techniques can be allocated to one of the categories outlined in (74). Some visualisations are highly specialised on the application domain and may therefore not fit well in one or more of the categories.

## Use of Animation

According to the authors of (40) animation has become popular in graphical user interfaces due to its engaging nature. Animation can be used to facilitate perception of changes in the data and data graphics (40). However, the authors of (40) also claim that animation has to be used with care as it may contribute to distraction if used inappropriately. Inappropriate use of animation could be for example animating irrelevant information or change that may grasp the users attention and thus misleads the users biological visual analysis system. This potential danger has resulted in research that aims to direct the usage of animation in general.

In this section we highlight three general kinds of animation for data and information visualisation that can be found in the literature and visual analytics systems. The kinds of animation are *animation of viewports*, *animated transition of graphs* and *animated time*. With the term viewport we refer to a typically rectangular viewing area that is of interest to the observer. *Animated viewpoints* are visualisations of the navigation space with respect to changes to the current viewport. These changes can happen for example through user interactions such as discussed in 'Interaction Techniques' section. For example filtering of data items, zooming into the display area, changing the level of detail etc. If these changes happen abruptly, then they may be too difficult to be comprehended by the observer. Animation can be used to allow the user to maintain an overview during interaction (70). The authors of (70) presented a generic model for the smooth animation of such changes that takes, amongst other aspects, the optimal animation velocity into account. This smooth viewport animation has been implemented and used for example by (14; 27). *Animated transitions of graphs* is the animation of switches from one data representation to another. For example the transition from a bar chart to a pie chart in order to see the relative percentages of the data represented in the bar chart. Animation of such transitions is intended to allow the user to identify elements across diverse representations (40). The authors of (40) have investigated animated transitions between statistical data representations and derived guidelines for the design of animated transitions, which they applied in their *DynaVis* visualisation system. With *animated time* we refer to animated changes over time in the data rather than changes on the viewpoint or the representation. The representation of time can be simply facilitated by using the display space, also known as static mapping, or through animation by using the physical time, also known as *dynamic mapping* (58). Both, static representation of time as well as dynamic representation of time are important

for visual analytics. For example the static representation of time allows the user to observe all available information and to compare the data with respect to the time, whereas dynamic mapping allows to observe the general development of the data over time (21).

However, the disadvantage of animation in general is that the data may simply be too complex to be perceived by the viewer (68), this is also the reason why the authors of (14) refrain from animating more than one view simultaneously.

## Available Visual Data Analysis Software Systems

Standard data analysis systems such as SPSS (43) and Weka (2) already provide a wide range of interactive visualisation techniques and data views. However, their visualisation techniques' aim is more for reporting and the interpretation of the data mining results and data statistics rather than for the actual analysis of the data and the interactive adjustment of the data mining models. Below we highlight some open source and some commercial data mining tools that allow to visually analyse data, some of which are tailored to a particular kind of data analysis application.

Bak et al (22) uses a combination of interactive visual analysis methods in order to effectively analyse multivariate datasets with demographic data. The authors use a Self-Organising Map (SOM) (50) to visualise each cluster as a radial parallel coordinate plot (69). Opacity bands (32) within this plot illustrate the variance within a cluster. The background colour coding is used to correlate the cluster with a target value. User interaction is facilitated at different stages of the analysis as well as within the visualisation (22). For example the user can modify the colour maps and their scaling and can choose different types of scaling such as logarithmic scaling. The visualisation also provides a 'mouse-over' which the user can utilise in order to identify individual cluster members and to display their data. The same techniques have also been used in order to analyse minority or ethnic residential patterns (56).

A complete implementation of an interactive graph visualisation system, called Coordinated Graph Visualisation (CGV) has been presented by (67), a test version can be obtained from (14).

Figure 4 shows a screenshot of the system. The main focus of the CGV system is on user interaction, addressing in particular the problem of how to interact with a graph that is too large to be displayed on a computer screen. CGV is based on hierarchy tree computations and graph macro-views that can be displayed in the available display space (16). Multiple views allow the user to explore, interact and access the data from different perspectives simultaneously. Figure 4 shows the basic graphical user interface of CGV. The main canvas of *main graph view* uses a node-link representation of a macro-view graph. In order to represent the graph hierarchy CGV uses several coordinated views. The *hierarchy view* (on the top of Figure 4) uses a superimposed polyline in order to visualise the current antichain and a colour coding for illustrating selected node attributes.

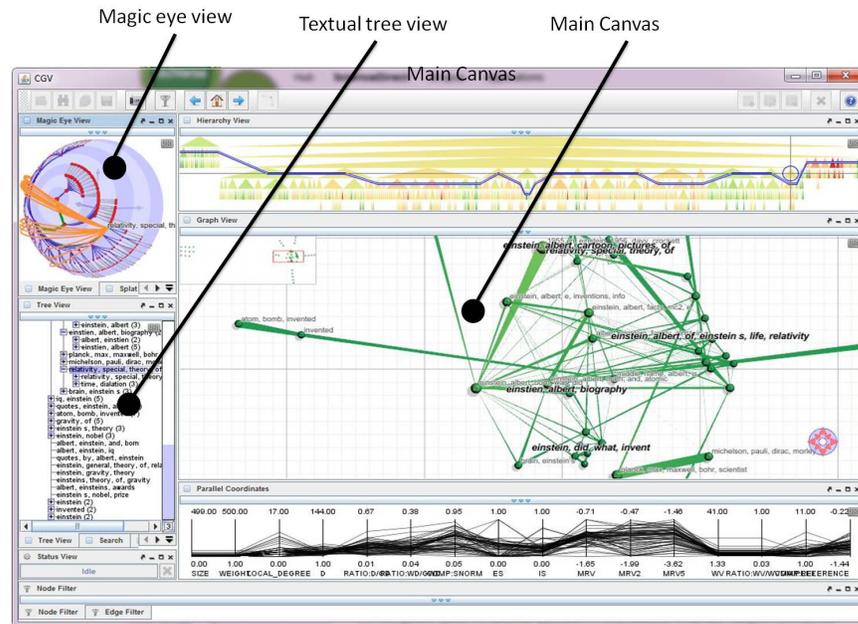


Figure 4: The graphical user interface of the CGV system.

The complementary *textual tree view* and the *magic eye view* are displayed enlarged in Figure 5. The *textual tree view* (on the left hand side of Figure 4 and on the top in Figure 5) is collapsible and complements the hierarchy view by providing textual labels. The textual tree view can also be magnified by a fisheye transformation. The *magic eye view* (on the top left corner of Figure 4 and in Figure 5) is based on the Walker Layout (25) and is extended by cross-edges amongst nodes. These cross edges are visualised using arcs spanning around the eye. However, there are further views in the CGV system, and we refer the interested reader to (67) for further details. CGV distinguishes between *coordinated* and *uncoordinated interactions*. Uncoordinated interactions are specific to the particular view, such as colour or font size etc., whereas coordinated interactions are changes to the global perspective of the data and hence are propagated to all views so that all views are consistent in their representation of the data. CGV supports basic user interactions such as zooming, panning, fisheye magnification, identifying objects (coordinated), locate objects (coordinated), lock/unlock & brushing the focus (coordinated), expand/collapse graph or tree structure (coordinated), visual parameter adjustment (uncoordinated). CGV implements filtering strategies that aim to filter out irrelevant nodes and thus enhance the clarity of the graph representation (67). Furthermore CGV implements view space and data navigation techniques, visual augmentation techniques and undo operations which are not discussed further here. However, a comprehensive discussion can be found in (67).

The NFloVis system (54) is a visual analytics system for the visualisation of NetFlow data (IP traffic information) in a computer network in order to detect network attacks.

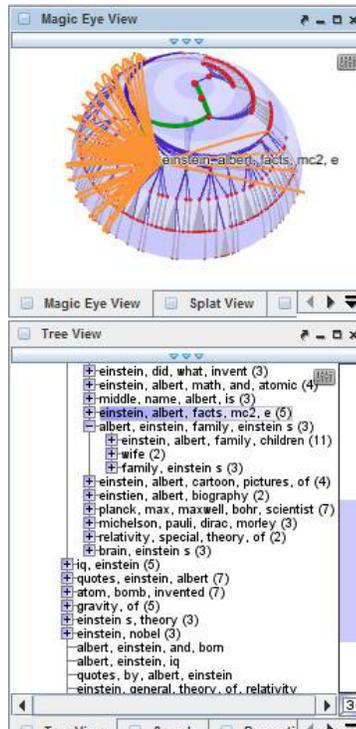


Figure 5: The textual tree view and magic eye view of the CGV system.

NFlowVis uses two principal visualisation techniques, *home centric flow visualization* and *graph-based flow visualisation*. In the home centric approach the local hosts that are related to attacking hosts are visualised in a TreeMap (63) and attacker hosts are placed at the borders. Flows between attackers and local hosts are visualised using splines. The colour of the local hosts and their size and the colour of the splines can be used to represent various properties, such as packets or bytes transferred (31). Thresholds can be used to hide splines with a low traffic and highlight splines with a high traffic to the attackers as shown in the screenshot of NFlowVis in Figure 6.

Some commercial visual analytics tools exist, such as *Miner3D* (15) for 3D data visualisation of multi-dimensional data, *Panopticon* (12) which allows to visualise real time streams of data and *SAS Visual Analytics* (61). Whereas SAS Visual Analytics also emphasises on the system’s computational scalability to large data volumes, which is achieved through in-memory processing. Another tool for visual data analytics specialised for large quantities of data is *Spotfire* (64). Also long established data analytics tools such as Matlab (45) have recognised the importance of data visualisation and model-based visualisations components such as in Matlabs Simulink extension (46).

As mentioned before, some tools such as *Panopticon* are tailored for the visualisation of data streams. the following section addresses the challenges in visualising data streams

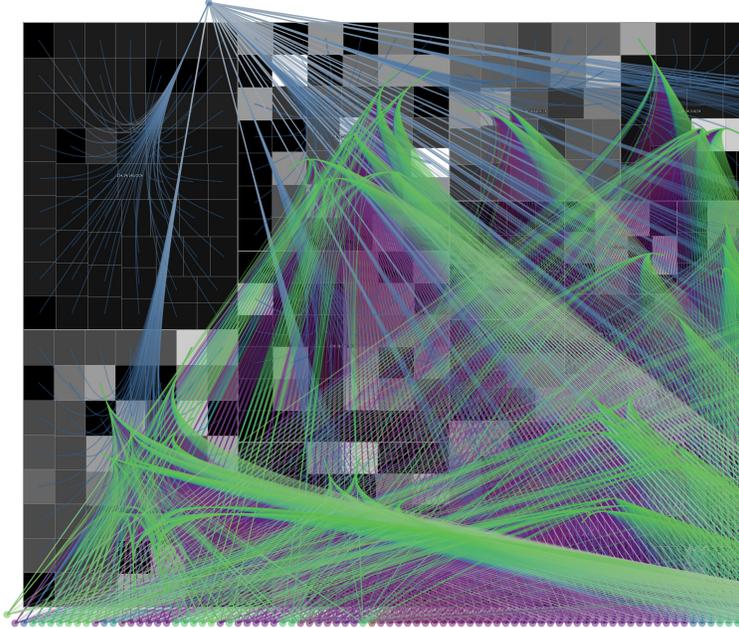


Figure 6: The figure shows a brute force SSH attacks from the Internet to computers located at the University of Konstanz (31; 54). The figure has been reproduced with accordance to the Creative Commons Attribution-ShareAlike 3.0 Unported License ([http://creativecommons.org/licenses/by-sa/3.0/deed.en\\_US](http://creativecommons.org/licenses/by-sa/3.0/deed.en_US)).

and data stream analysis techniques.

## Visual Data Mining for Data Streams

Mining data streams has attracted a big deal of attention over the last decade. A large number of techniques have been proposed to address the problem of analysing high-rate streaming data in real-time, such as incremental learning algorithms (24). Gaber et al (36) have provided a review of clustering, classification and frequent pattern mining techniques adopted for mining of streaming data. In a more recent review, a taxonomy of notable techniques in the area has been provided by Gaber in (34). This taxonomy divided stream mining techniques to: (a) two phase techniques that start with an on-line stage, followed by an offline one; (b) Hoeffding bound techniques that provide statistical guarantee on the approximation of the data mining output; (c) symbolic approximation that proved to be the state-of-the-art approach to time series analysis; and (d) granularity-based techniques that provide a framework for resource-adaptive data stream mining.

Limited work has been proposed for visualisation of data stream mining results. This is due to the dynamic nature of the results, and also to the potential large knowledge structures produced by the process (number of clusters, number of levels in a decision tree, etc.). Massive Online Analysis (MOA) (23) is a comprehensive data stream mining tool that provides, among its functionalities, dynamic visualisation of data stream clustering.

Visualisation of the results of data stream mining becomes more challenging when the process run on mobile devices with small screen real-estate like smartphones. Mobile data stream mining has been well researched over the last few years (52). Open Mobile Miner is a stream mining tool that is tailored to function on mobile platforms (38). The screen of the mobile device can get cluttered rapidly when trying to visualise the results of a data stream mining process, running onboard the device. Addressing the problem adopting clutter reduction techniques has been reviewed in (30). The reviewed techniques addressed the problem in a general way, not for mobile devices specifically. Ellis and Dix (30) have categorised the clutter reduction techniques into three categories: (a) representation change of data items; (b) distance change among data items; and (c) animation when a temporal dimension of the data exists. Although these approaches deal with the clutter effectively in different cases, they are not self-adaptive.

Adaptability to screen clutter for mobile devices has been proposed by Gaber et al (34). The approach is based on a theoretical underpinning that balances perception, clutter, and amount of information of the visual image, having two observations: (a) the more information, the greater the clutter of the visualised image; (b) the higher the clutter, the less the perception of the image. With the objective of increasing the information and reducing the clutter to enhance the perception, Gaber et al (34) have developed Adaptive Clutter Reduction (*ACR*) (35). *ACR* is a generic theory that states an optimisation function that works on maximising both, the information and perception while reducing the amount of clutter.

*ACR* has been applied to data stream clustering on the smartphone. The technique has been coined Clutter-Aware Cluster Visualisation (*CACV*). The process is divided into two steps. The first is applying the data stream clustering technique (RA-Cluster) (33). The visualisation of the evolving clusters is done in the second step. Applying the *ACR* theory, all clusters are presented to satisfy two requirements: (a) maximum allowable screen coverage by the clusters, and (b) maximum allowable percentage of cluster overlapping. Accordingly, a heuristic based technique is applied. One of four levels of visualisation is applied at any point in time to satisfy the two requirements. The levels are ranked according to the amount of information. Thus, the higher the level, the more information is presented on the screen. When a clutter situation occurs by having at least one of the two requirements not satisfied, the level of visualisation changes to the lower ranked in the list. These levels are (a) normal: clusters are presented on the screen having the size representing the number of points in the cluster, (b) scaling: clusters are scaled down in an attempt to satisfy the two requirements, (c) colouring/shading: all clusters would have the same size, having the darker ones represent those clusters with larger number of points, (d) active mode: only active clusters are presented on the screen. It can be observed that each level of visualisation rep-

resents the level of information. Thus, CACV satisfies the requirements of the *ACR* theory, by keeping the highest information and perception level, while minimising the screen clutter.

Visualisation of data stream mining results is still in its infancy state. *ACR* theory represents an attempt to generalise the solution for small screen real-estate of smartphones. Nevertheless, the approach is applicable for larger screens. However, more work on the application of *ACR* to other data mining techniques is needed.

## **Visual Data Mining Applications in the Chemical Process Industry**

The application of Visual Analytics to commercial as well as to scientific problems is becoming popular due to the growing availability of commercial as well as scientific Visual Analytics tools as highlighted in the '*Available Visual Data Analysis Software Systems*' section. This section discusses the usage of Visual Data Analytics from the perspective of the chemical process industry as an example. However, further applications of Visual Data Analytics are highlighted in the '*Further Applications of Visual Data Analysis*' section.

As an example in this article we use the case of Evonik Industries AG, which is one of the biggest global players in the chemical process industry. As of 2012 Evonik Industries AG employed roughly 33,000 employees and operates production plants in 24 countries. This and more information on Evonik Industries AG can be found at <http://evonik.com>.

Evonik Industries AG, generates massive amounts of data about production processes and in laboratory scale research. Typical applications in this industry that collect data are Process Control Systems (PCSs), Process Information Systems (PIMSs) and Laboratory Information Management Systems (LIMSs). The collection of these systems is used to define Key Performance Indicators (KPIs). There are two main user groups of these systems, the *production* and the *research and technical support*. The aforementioned systems are generally used for reporting and for the improvement of the production in terms of reducing cost and improving the product quality.

For doing so, the production uses the KPIs but also aims to predict the future product quality using data mining techniques. However, within the production group the view on Visual Data Mining differs between plant operators and plant managers. For example the actual plant operator desires detailed information about the actual process' state of a single active plant; the operator uses this information for supervising the current production. The plant manager desires a less detailed overview of the production but for several plants and over a longer period of time. This condensed information for the manager often needs to be displayed in so called information dashboards. Information dashboards are easy to read condensed data representations often only showing the current status and historic and future trends. The manager uses this information for reporting and improvements of the production processes (17). Yet the data for both,

the plant operators and the plant managers, is the same. Hence, Visual Data Analytics needs to support filtering and other interaction techniques accordingly. The *Examples of Visualisations in the Production* section gives examples for the use of Visual Data Analytics in Evonik Industries AG within the production user group.

The research user group's goal is to improve the products' recipes in order to improve the production quality and lower the cost of raw material. This group aims to automatise their laboratory work by using High Throughput Screening (HTS) and robotics in order to conduct a multitude of tests in a short period of time, examples can be found here (51; 39). In order to analyse the experimental data this group uses Information Visualisation and Visual Data Analytics techniques. The *'Examples of Visualisations in Research'* section gives examples for the use of Visual Data Analytics in Evonik Industries AG within the production user group.

## **Examples of Visualisations in the Production**

The visualisation of real time data during a production process aims to identify the process's 'capability index', which is a measure used to estimate the optimal production capacity. There is a balance between changing production parameters (which may destabilise the process) and keeping the process stable. Statistical information and data visualisation helps to decide about interventions. At Evonik Industries AG an application called ChemSPF (71), an in-house development, is used for the statistical evaluation in process analysis and protection of process capability. The ChemSPF-chart and histogram in Figure 7 show some of the visualisations used to display quality attributes of an ongoing production. The ChemSPF-chart is on the bottom half of Figure 7, it also displays the mean value (MV), upper (USL) and lower limits (LSL) shown for triggering a warning, as well as upper and lower limits for triggering interventions (UAL and LAL). Please note that the acronyms used in Figure 7 are based on the German language and hence the acronym expansions are omitted here. Also the name ChemSPF charts comes from a German expression and is also known as Statistical Process Control (SPC) charts in the English language.

These are just two examples of standard visualisations used in the production. However, there are many more including text based information. What is important to note is that the charts in Figure 7 display only one variable. However, in a production process there may be hundreds of variables with different value ranges which have to be observed in real time. Hence, these charts have to be redrawn in adequate time intervals and need to convey large amounts of information simultaneously.

## **Examples of Visualisations in Research**

Whereas the production group uses mostly basic visualisations, the research group needs to observe as many variables of historic productions or lab scale productions within the same visualisation. The task for the researcher is to extract as much information as possible from the data. Evonik Industries AG uses an application that makes

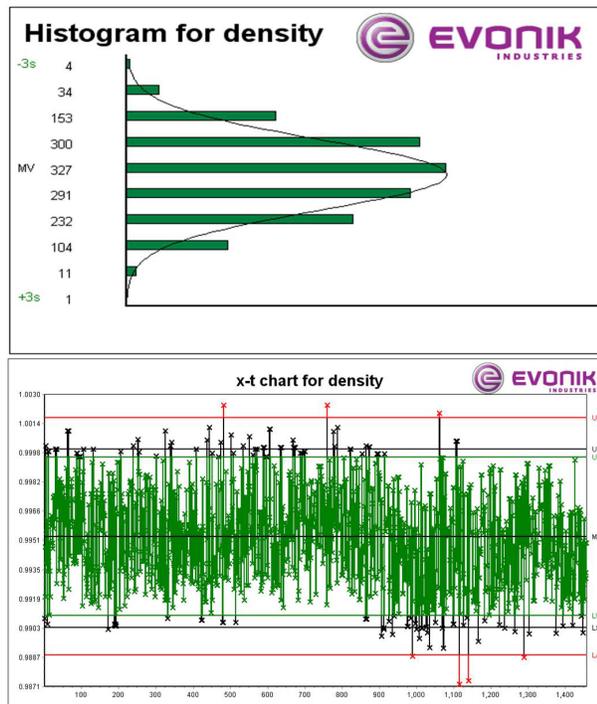


Figure 7: Charts used by Evonik Industries AG to display quality attributes during the production process. The chart on the bottom half of the figure is called a ChemSPF-chart and the chart on the top half of the figure is a histogram.

use of Visual Data Analytics techniques, which is used by researchers for analysing large quantities of data. The application is called KinFit (57) and is an in-house development. One task of Visual Analytics in research at Evonik Industries AG is to find correlations between variables. Therefore KinFit gives a data overview by showing each dimensions plotted against each other, which is also known as scatter matrix in other data mining tools such as WEKA (2). Within Evonik Industries AG the Visual Analytics techniques in KinFit are used for exploring production data and for general research purposes. Information Visualisations in KinFit are also used for reporting.

Many more interactive visualisation approaches are used. For example in order to visualise multiple dimensions (variables) in the same chart, multi dimensional charts such as displayed in Figure 8 are frequently used. The user can change the encoding of the data elements such as colours, symbols and the symbol's size as well as moving, zooming or rotating the graph. KinFit also allows to animate the graph in order to show a sixth dimension such as time.

A polar or radar chart such as the one displayed in Figure 9, also called spider chart, shows in each spoke one dimension. Each line is representing one data record. This chart can be used for detecting outliers and clusters. The user can interact with the

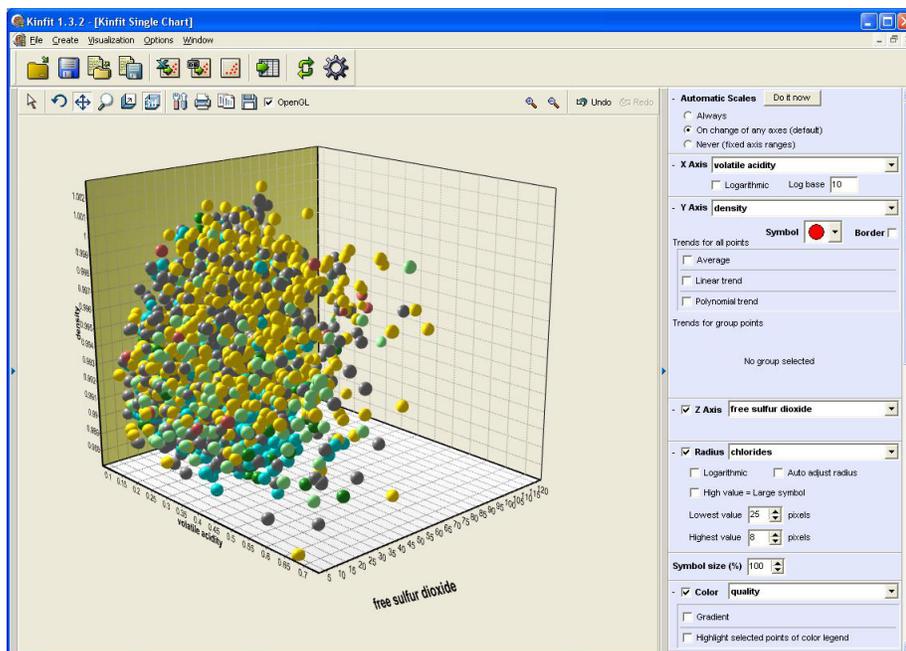


Figure 8: Multi dimensional representation of five variables using KinFit.

graph parameters and set for each dimension an optimal value and a weight. The application then gives the range of the most optimal records as displayed in Figure 9.

However, one challenge of Visual Analytics that remains to be solved in the chemical process industry is to convey simultaneously thousands of variables with different and changing values ranges in real time.

## Further Applications of Visual Data Analysis

The last two subsections showed that Visual Data Analytics is one of the emerging technologies for data analysis in the chemical process industry. Also a range of applications have been mentioned throughout the article, especially in the '*Available Visual Data Analysis Software Systems*' section. However, further applications exist. For example the authors of (42) propose a two stage visualisation approach for fraud detection in stock market trading. In the first stage they use 3D tree maps to observe real-time stock market performance and detect unusual patterns; and in the second stage they use social network visualisation to analyse the behaviour of the suspected pattern in order to detect imminent attacks. In (59) the authors presented a Visual Analytics tool that visualises the aggregate risk alongside a traditional wealth-time plot. The system is aimed at non-expert users to help with personal finance decisions. The system has been used

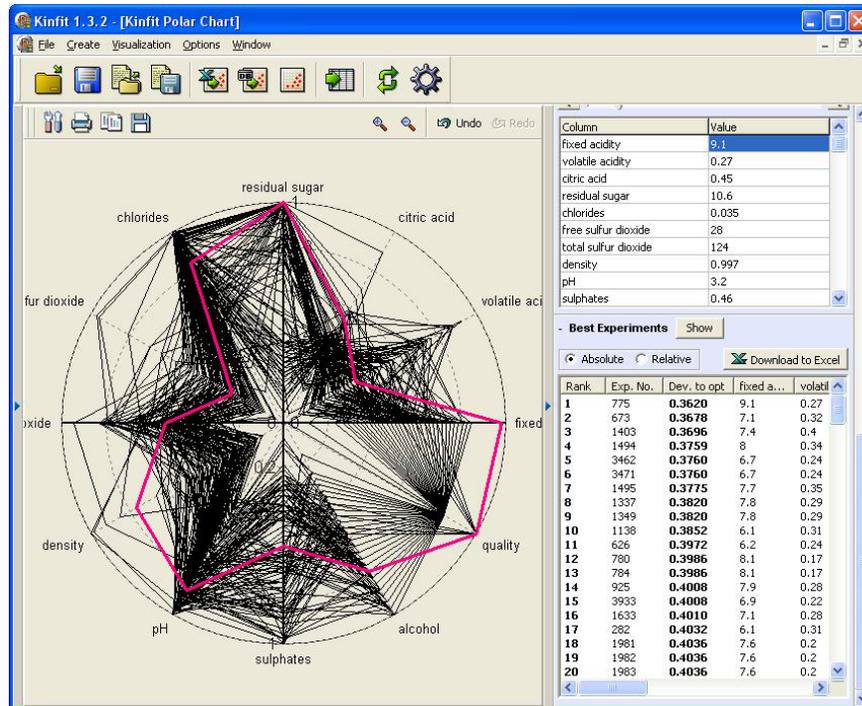


Figure 9: KinFit's polar chart implementation for displaying data with multiple dimensions, typically used for identifying optimal production process states.

in a laboratory environment by 27 volunteers and been evaluated using several economics methods. For example the results shows that the users gain greater returns for similar levels of risk. Also it shows that users of the system explore and learn more than the control group, which is reflected in a higher number of investment modifications of system users. Also 23% of the users of the system claimed that they feel more confident in understanding financial planning whereas 0 % of the control group made this claim. In general the results show that the system improves the decision making process. The authors of (28) applied Visual Analytics on multidimensional gene expression datasets. Their approach is to visualise the gene expression conjoined with statistical data retrieved from the original data. (60) uses visual analysis techniques in the physics arena in order to gain insights into simulations of magnetically confined burning plasmas. The authors of (49) have developed a Visual Analytics system that has been used in a variety of tasks in the context of analysing hyperspectral images of historical documents.

## Concluding Discussion and Major Challenges

As discussed in the previous sections of this paper Visual Analytics as a field of study is only just emerging. We have attempted to explain that this new field is a combination of interactive visualisation and advanced data and predictive analytics which includes building, verifying and using predictive models. As shown in Figures 1 and 2 it is not only the visualisation of the data and its exploratory analysis, the process which is often mistakenly given the name of Visual Analytics, but also an interactive method available for semi-automated clustering, classification, prediction, dimensionality reduction, information retrieval, feature extraction etc, which leads to knowledge and actionable decision generation.

In the process of such a Visual Analytics session one would expect the underlying predictive models to be searching for and exploiting user feedback in many different forms. Such feedback could take a form of extracting user preferences by monitoring what is used and what is focused on, which in turn could be translated into modified optimisation criteria by the underlying models to adapt to the user's current needs. We would also expect that the users should be provided with powerful wizard driven ways of exploiting and building data processing methods without a need to understand or set any of the technical parameters of such advanced methods. This represents the model building and model visualisation interactive loop in Figures 1 and 2.

The visualisation of extracted knowledge in turn drives further refinements. This focus on the data or how the data or the existing models should be transformed closes the loop which can be repeated a number of times. This process becomes even more challenging and interesting when the data arrives continuously (i.e. streaming) and the predictive models must operate in a (semi-)automated manner. Visualisation of the changing knowledge that is brought to the attention of the user and suggestions of what may be interesting in the recent data are then also a part of the picture, a picture that may be changing very quickly.

There is a growing number of visual data analysis software tools like the ones highlighted in the '*Available Visual Data Analysis Software Systems*' section and various components required for building of visual analytics systems and tools are also being developed and discussed in related fields as highlighted in the '*Typical Concepts and Components of Visual Data Analytics Systems*' section, but what remains a formidable challenge is to bring all of the components together in robust tools that can be used by users with a wide range of technical abilities and know-how. As illustrated in a snapshot of visualisation tools that are used by one of our partners in the process industry (see Section '*Visual Data Mining Applications in the Chemical Process Industry*') we are far from such visual analytics scenarios as the current practises very often end on interactive visualisation for which the tools are available. It is the role of the data analysts to explore the data and look for things that may be interesting, unusual, dangerous, promising. We are still very heavily relying on the creativity and curiosity of humans who in addition must have quite formidable technical abilities and knowledge of data and predictive analytics tools to realise their full potential.

In developing a next generation of Visual Analytics systems one would hope to relax (if not remove completely) this need for the user's deep knowledge of the predictive analytics methods and enhance the ability of Visual Analytics systems so they in part can become more autonomous and 'creative' and 'curious' in looking for things that are interesting in some sense. To reach such a state a lot of research and development effort is still needed and one of this publication's goals is to bring this fact to the attention of wider communities with keen interest in what is emerging as the Visual Analytics field.

### Supplementary Information

The research leading to these results has received funding from the European Commission within the Marie Curie Industry and Academia Partnerships & Pathways (IAPP) programme under grant agreement n° 251617.

### References

- [1] John Gantz and David Reinsel. The digital universe decade, are you ready? *IDC*, 2009(May):1–16, 2010.
- [2] Ian H Witten and Frank Eibe. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, second edition, 2005.
- [3] B McClean, C Hawkins, A Spagna, M Lattanzi, B Lasker, H Jenkner, and R White. New horizons from multi-wavelength sky surveys. *Proceedings of the 179th Symposium of the International Astronomical Union held in Baltimore*, 1998.
- [4] Sloan digital sky survey, accessed February 2013. <http://www.sdss.org/>.
- [5] Human genome project, accessed February 2013. [http://www.ornl.gov/sci/techresources/Human\\_Genome/home.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml).
- [6] F. Stahl, M.M. Gaber, M. Bramer. Scaling up Data Mining Techniques to Large Datasets Using Parallel and Distributed Processing In Rausch, P. Sheta, A. F. and Ayes, A. (eds.) *Business Intelligence and Performance Management. Advanced Information and Knowledge Processing*. Springer London, pp. 243-259.
- [7] Zaixian Xie. *Exploratory Visualization of Data Pattern Changes in Multivariate Data Streams*. PhD thesis, WORCESTER POLYTECHNIC INSTITUTE, 2011.
- [8] Daniel A. Keim, Florian Mansmann, Jorn Schneidewind, and Hartmut Ziegler. Challenges in visual data analysis. In *Proceedings of the conference on Information Visualization, IV '06*, pages 9–16, Washington, DC, USA, 2006. IEEE Computer Society.
- [9] S.K. Card, J.D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann Publishers, 1999.

- [10] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, pages 336–343, Washington, DC, USA, 1996. IEEE Computer Society.
- [11] James J. Thomas and Kristin A. Cook (Eds). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.
- [12] Panopticon real-time visual data analysis, accessed February 2013. <http://www.panopticon.com/>.
- [13] *Computer Graphics: Principles & Practice In C, 2/E*. Pearson Education, 1996.
- [14] CGV - coordinated graph visualization, accessed February 2013. <http://www.informatik.uni-rostock.de/~ct/software/CGV/CGV.html>.
- [15] The miner3d data visualisation software, accessed February 2013. <http://www.miner3d.com/>.
- [16] James Abello, Frank van Ham, and Neeraj Krishnan. Ask-graphview: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):669–676, September 2006.
- [17] M. Adams and W.K. PSI AG. Din en 62264. die neue norm zur interoperabilitaet von produktion und unternehmensfuehrung–teil 1., Automatisierungstechnische Praxis, Vol. (49) 5.2007, S. 52-57, 2007. 2007.
- [18] KNIME.com AG. Eclipse rich client platform, accessed February 2013. <http://www.eclipse.org/home/categories/rcp.php>.
- [19] KNIME.com AG. Knime - professional open-source software, accessed February 2013. <http://www.knime.org/>.
- [20] Christopher Ahlberg, Christopher Williamson, and Ben Shneiderman. Dynamic queries for information exploration: An implementation and evaluation. pages 619–626. ACM Press, 1992.
- [21] Wolfgang Aigner, Silvia Miksch, Wolfgang Mueller, Heidrun Schumann, and Christian Tominski. Visualizing time-oriented data-a systematic view. *Computers and Graphics*, 31(3):401 – 409, 2007.
- [22] Peter Bak, Itzhak Omer, and Tobias Schreck. Visual Analytics of Urban Environments using High-Resolution Geographic Data. In *M. Painho et al. (eds.) Geospatial Thinking, Lecture Notes in Geoinformation and Cartography*, pages 25–42. Springer Berlin, 2010.
- [23] Albert Bifet, Geoff Holmes, Richard Kirkby, Bernhard Pfahringer, and Jansen T. Seidl T Kranen P., Kremer H. Moa: Massive online analysis. *Journal of Machine Learning Research (JMLR)*, 11:1601–1604, 2010.

- [24] A. Bouchachia, Bogdan Gabrys, and Z. Sahel. Overview of some incremental learning algorithms. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2007): Proceedings*, pages 1–6. London, UK, July 2007, 2007.
- [25] Christoph Buchheim, Michael Jünger, and Sebastian Leipert. Improving walker's algorithm to run in linear time. In *Revised Papers from the 10th International Symposium on Graph Drawing, GD '02*, pages 344–353, London, UK, UK, 2002. Springer-Verlag.
- [26] C. Chen. *Information Visualization: Beyond the Horizon*. Computer Science. Springer, 2006.
- [27] Andy Cockburn, Joshua Savage, and Andrew Wallace. Tuning and testing scrolling interfaces that automatically zoom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05*, pages 71–80. ACM, 2005.
- [28] J. Dietzsch, J. Heinrich, K. Nieselt, and D. Bartz. Spray: A visual analytics approach for gene expression data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 179–186, oct. 2009.
- [29] A. Dix. *Human-Computer Interaction*. Pearson/Prentice-Hall, 2004.
- [30] Geoffrey Ellis and Alan Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, November 2007.
- [31] Fabian Fischer, Florian Mansmann, Daniel A. Keim, Stephan Pietzko, and Marcel Waldvogel. Large-scale network monitoring for visual analysis of attacks. In *Proceedings of the 5th international workshop on Visualization for Computer Security, VizSec '08*, pages 111–118, Berlin, Heidelberg, 2008. Springer-Verlag.
- [32] Ying-Huey Fua, Matthew O. Ward, and Elke A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the conference on Visualization '99: celebrating ten years, VIS '99*, pages 43–50, Los Alamitos, CA, USA, 1999. IEEE Computer Society Press.
- [33] M.M. Gaber and P.S. Yu. A holistic approach for resource-aware adaptive data stream mining. *New Generation Computing*, 25(1):95–115, 2006.
- [34] Mohamed Medhat Gaber. Advances in data stream mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):79–85, 2012.
- [35] M.M. Gaber, S. Krishnaswamy, B. Gillick, H. AlTaiar, N. Nicoloudis, J. Liono, A.B. Zaslavsky. Interactive self-adaptive clutter-aware visualisation for mobile data mining. *Journal of Computer and System Sciences*, 79(3): 369-382 (2013).
- [36] Mohamed Medhat Gaber, Arkady Zaslavsky, and Shonali Krishnaswamy. Mining data streams: a review. *ACM Sigmod Record*, 34(2):18–26, 2005.

- [37] Macrofocus GmbH. Infoscope, accessed February 2013. <http://www.macrofocus.com/public/products/infoscope/>
- [38] P. Haghghi, S. Krishnaswamy, A. Zaslavsky, M. Gaber, A. Sinha, and B. Gillick. Open mobile miner: a toolkit for building situation-aware data mining applications. *Journal of Organizational Computing and Electronic Commerce*, accepted to appear.
- [39] C. Heddle and S.L. Mazaleyrat. Development of a screening platform for directed evolution using the reef coral fluorescent protein zsgreen as a solubility reporter. *Protein Engineering Design and Selection*, 20(7):327–337, 2007.
- [40] J. Heer and G.G. Robertson. Animated transitions in statistical data graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1240–1247, 2007.
- [41] Harry Hochheiser and Ben Shneiderman. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, March 2004.
- [42] Mao Lin Huang, Jie Liang, and Q.V. Nguyen. A visualization approach for frauds detection in financial market. In *Information Visualisation, 2009 13th International Conference*, pages 197–202, July 2009.
- [43] IBM. Spss software, accessed February 2013. <http://www-01.ibm.com/software/analytics/spss/>.
- [44] Google Inc. Google earth, accessed February 2013. <http://www.google.com/earth/index.html>.
- [45] Explore, model, and visualize data, accessed April 2013. <http://www.mathworks.co.uk/data-analysis/>.
- [46] Simulation and Model Based Design, accessed April 2013. <http://www.mathworks.co.uk/products/simulink/>.
- [47] D.A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann. *Mastering the Information Age - Solving Problems with Visual Analytics*, 2010.
- [48] Daniel A. Keim, Florian Mansmann, Jörn Schneidewind, Jim Thomas, and Hartmut Ziegler. Visual data mining. chapter Visual Analytics: Scope and Challenges, pages 76–90. Springer-Verlag, Berlin, Heidelberg, 2008.
- [49] Seon Joo Kim, Shaojie Zhuo, Fanbo Deng, Chi-Wing Fu, and Michael Brown. Interactive visualization of hyperspectral images of historical documents. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1441–1448, November 2010.
- [50] T. Kohonen. *Self-Organizing Maps*. Springer Series in Information Sciences. Springer, 2001.

- [51] Tanuja Koppal. Ask the expert setting up high-throughput screening laboratory, 2012.
- [52] S. Krishnaswamy, J. Gama, and M.M. Gaber. Advances in data stream mining for mobile and ubiquitous environments. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 2607–2608. ACM, 2011.
- [53] Jens Kruger, Jens Schneider, and Rudiger Westermann. Clearview: An interactive context preserving hotspot visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):941–948, 2006.
- [54] Florian Mansmann, Fabian Fischer, Daniel A. Keim, and Stephen C. North. Visual support for analyzing network traffic and intrusion detection events using treemap and graph representations. In *Proceedings of the Symposium on Computer Human Interaction for the Management of Information Technology, CHI+IT '09*, pages 3:19–3:28, New York, NY, USA, 2009. ACM.
- [55] Malgorzata Migut and Marcel Worring. Visual exploration of classification models for various data types in risk assessment. *Information Visualization*, 11(3):237–251, 2012.
- [56] I. Omer, P. Bak, and T. Schreck. Using space-time visual analytic methods for exploring the dynamics of ethnic groups’ residential patterns. *Taylor & Francis International Journal of Geographical Information Science*, 24(10):1481–1496, 2010.
- [57] D Porwol. User-manual kinfit 2.0. Manual, Evonik Industries AG, 2012.
- [58] A. Rind, B. Neubauer, W. Aigner, S. Miksch, K. Matkovic, and G. Santucci. Static and dynamic visual mappings to explore bivariate data across time. *EuroVA 2012 Poster Proceedings*, page 3, 2012.
- [59] S. Rudolph, A. Savikhin, and D.S. Ebert. Finvis: Applied visual analytics for personal financial planning. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 195–202, oct. 2009.
- [60] Allen Sanderson, Guoning Chen, Xavier Tricoche, David Pugmire, Scott Kruger, and Joshua Breslau. Analysis of recurrent patterns in toroidal magnetic fields. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1431–1440, November 2010.
- [61] SAS. Sas visual analytics. accessed February 2013. <http://www.sas.com/technologies/bi/visual-analytics.html>
- [62] SAS. Sas enterprise guide, accessed February 2013. [http://www.sas.com/technologies/bi/query\\_reporting/guide/](http://www.sas.com/technologies/bi/query_reporting/guide/).
- [63] Ben Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.*, 11(1):92–99, January 1992.

- [64] TIBCO Software. Spotfire, accessed February 2013. <http://spotfire.tibco.com/>.
- [65] R. Spence. *Information visualization: design for interaction*. Pearson/Prentice Hall, 2007.
- [66] Jim Thomas and Joe Kielman. Challenges for visual analytics. *Information Visualization*, 8(4):309–314, 2009.
- [67] Christian Tominski, James Abello, and Heidrun Schumann. CGV-an interactive graph visualization system. *Computers and Graphics*, 33(6):660 – 678, 2009.
- [68] B. Tversky, J.B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.
- [69] Tran Van Long and Lars Linsen. Multiclustertree: Interactive visual exploration of hierarchical clusters in multidimensional multivariate data. *Computer Graphics Forum*, 28(3):823–830, 2009.
- [70] J.J. van Wijk and W.A.A. Nuij. A model for smooth viewing and navigation of large 2d information spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 10(4):447–458, 2004.
- [71] G Vornholt. User-manual chemspf 4.0. Manual, Evonik Industries AG, 2008.
- [72] C. Ware. *Information Visualization: Perception for Design*. Interactive Technologies. Elsevier Science, 2012.
- [73] Ji S. Yi, Rachel Melton, John Stasko, and Julie A. Jacko. Dust Magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(4):239–256, 2005.
- [74] Ji Soo Yi, Youn ah Kang, John Stasko, and Julie Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, November 2007.