INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team gravite

Graph Visualization and Interactive Exploration

Bordeaux - Sud-Ouest

Theme : Knowledge and Data Representation and Management

Table of contents

GRAVITÉ is a joint project with CNRS (UMR 5800 LaBRI) and Université de Bordeaux. The team has been initiated on April 2007 and was officially created as a project in January 2009.

1. Team

Faculty Members

David Auber [Associate Professor, Université Bordeaux 1]
Romain Bourqui [Associate Professor, IUT/Université Bordeaux 1]
Maylis Delest [Professor, Université Bordeaux 1, HdR]
Guy Melançon [Team leader, Professor, Université Bordeaux 1, HdR]
Bruno Pinaud [Associate Professor, Université Bordeaux 1]

Technical Staff

Jonathan Dubois [(Non permanent staff) Research Engineer, Université Bordeaux 1]
Ludwig Fiolka [(Non permanent staff) Research Engineer, Université Bordeaux 1]
Charles Huet [(Non permanent staff) Research Engineer, Université Bordeaux 1]
Patrick Mary [Research Engineer, Université Bordeaux 1]
Morgan Mathiaut [(Non permanent staff) Research Engineer, Inria]

PhD Students

Frédéric Gilbert [Université Bordeaux 1, MENRT scholarship]
Antoine Lambert [Université Bordeaux 1, CORDI-R]
Maurin Nadal [Université Bordeaux 1, MENRT scholarship]
François Queyroi [Université Bordeaux 1, MENRT scholarship]
Arnaud Sallaberry [Université Bordeaux 1, CIFRE scholarship]
Paolo Simonetto [Université Bordeaux 1, MENRT scholarship]
Trung Tien Phan Quang [Université Bordeaux 1, MENRT scholarship]
Faraz Zaidi [Université Bordeaux 1, SFERE scholarship]

Post-Doctoral Fellow

Daniel Archambault [(Non permanent staff) Post-Doc, INRIA]

Administrative Assistant

Alice Rivière [Secretary (SAR) Inria]

2. Overall Objectives

2.1. Introduction

A recent U.S. report to the funding agencies NIH and NSF [36] provides strong arguments in favor of the development of visualization as a research field.

> "Visualization is indispensable to the solution of complex problems in every sector, from traditional medical, science and engineering domains to such key areas as financial markets, national security, and public health. Advances in visualization enable researchers to analyze and understand unprecedented amounts of experimental, simulated, and observational data and through this understanding to address problems previously deemed intractable or beyond imagination."
>
> [*from the Executive summary of* [36]].

Visualization is becoming a critical issue in a number of areas. It is more and more seen as a complementary approach - not only as a tool - for data retrieval and exploration or strategic and competitive watch, for instance, where the structure of data is typically unknown and must be discovered and understood. It is also a reliable approach when navigating "already known" and structured information space such as large ontologies. Information Visualization usefully combines with Scientific Visualization to help expert users explore and understand complex models underlying large scale simulations.

In response to this challenge, GRAVITÉ aims at designing interactive visualization methods and tools to analyze and mine large datasets. Our emphasis is on the visualization of graph structures to help users gain insights from large datasets and large-scale simulations, to understand the data and/or the underlying model, and ultimately, to identify intrinsic properties or emergent phenomenon.

More than just being able to deal with large volume and inhomogeneous data, we are required to deal with constant changes in data, possibly making it ambiguous and uncertain. In the context of graph visualization, the challenge we face is thus to design methods and tools:

- to deal with large and dynamically changing graphs;
- to visually identify salient properties in changing substructures;
- identify the multiscale nature of data;
- to produce visual cues helping the user to track such changes in either one three different situations where dynamic graphs occur.

A clear priority for our research is to address the needs of expert users faced with interpreting dynamic data. Our agenda is based on the needs expressed by our research community [71], [36], [38]. Our collaborations with experts of other scientific fields as well as with industry contribute to the overall organization of this research agenda and serve a twofold objective:

- to build theoretical knowledge relevant to information visualization and visual analytics, and develop a sound methodology for graph visualization and navigation;
- to target transfer opportunities favoring the adoption of our ideas and technology by other scientific communities and by the industry.

2.2. Highlights

- The team has succesfully organized the Eurovis 2010 Eurographics/IEEE Symposium on Visualization between June 9-11 2010.
- Faraz Zaidi has been nominated for the best application paper award at the International French Speaking Conference EGC2010 and for the best paper award at the Industrial Conference on Data Mining 2010 (http://www.data-mining-forum.de/paper_award_2010.php).
- David Auber has obtained a European Project (RAISME project), for 2 years.
- Romain Bourqui and Bruno Pinaud have successfully applied to the ANR JCJC 2010 program for a 4 years project on dynamic graphs (EVIDEN project).

3. Scientific Foundations

3.1. Introduction

The visualization of data or information with the help of a computer, most often referred to as "Information Visualization", originated from ideas in fields as diverse as computer graphics, computer-human interaction, cognitive psychology, semiotics, graphics design, cartography and graphical arts. One of its early motivations was to help users explore and analyze large quantities of data by developing software tools exploiting human visual capabilities according to [90], 40% of our cortical activities are dedicated to processing visual signals.

The design of new visualization methods and tools becomes even more necessary with the continuously increasing volume of available data, which poses a problem that obviously cannot be solved by relying solely on the increase of CPU power. According to the « How much information » project developed at Berkeley, one exabyte of data (1 million terabytes) was produced in 2001, with 99,997% being exclusively available digitally (see [Keim 2001]). In 2003, that quantity seen as individual data production corresponded to 800 megabytes per person in one year on the whole planet [82]. This abundance of information of course raises many questions and problems to solve. A number of research fields now contribute in their own way to the design of methods and tools to exploit this richness of information, among which visual approaches experience growing success.

3.2. Visual Graph Mining

Visually mining data requires astutely combining data analysis with visual graphics and interaction. Mining itself draws not only on statistics but in a rather astute mixture of mathematical rigor and heuristic procedures. As David Hand puts it [67], [66] :

> "To many, the essence of data mining is the possibility of serendipitous discovery of unsuspected but valuable information. This means the process is essentially exploratory."

From Hand's perspective, we see that visualization has much to share with data mining because visualization often comes as an aid to exploratory analysis. The analysis task we are concerned with however differs from that conducted by dataminers, in that we seek to be able to produce readable and interactive visualizations rather than coming up with reasonable, arguable and final conclusions on the data. The perspective to adopt is a combination of (semi) automated data processing together with human analytical and perceptual capabilities. Although relying on technology, the analysis task remains in total control of the human user. The NVAC research agenda [38] clearly states:

> "[The] analysis process requires human judgment to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information [...]"

later calling for the development of

> [...] visually based methods to support the entire analytic reasoning process, [...].

That is, in ideal cases the visualization should be designed in order not only to assist the analysis but to also actively contribute to its progress. Visualization thus appears as a multi-disciplinary field embracing a large spectrum of competencies. This partly comes from the need to cover all processes involved in the so-called "Visualization pipeline" as depicted here:

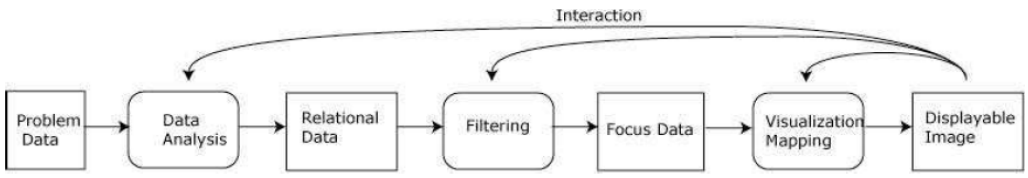

Figure 1. Visualization pipeline (adapted from [95]).

A decade ago, Ben Shneiderman[1] - who definitely helped Information Visualization to gain scientific visibility - suggested that visualization scenarios should obey his now celebrated mantra "Overview first, zoom and filter, then details on demand" [86]. The pipeline is coherent with Shneiderman's mantra which actually provides

[1] Ben Shneiderman is professor in the Department of Computer Science, and Founding Director (1983-2000) of the Human-Computer Interaction Laboratory at the University of Maryland (USA).

an excellent framework applying to almost any visualization environment. The back arrows correspond to the user interacting on the view, asking for details or zooming in on a particular subset of the data.

Daniel Keim has recently proposed a revised mantra, changing the focus towards data analysis[2]:

Analyse First - Show the Important - Zoom, Filter and Analyse Further - Details on Demand

Keim's mantra is closer to our perspective, merging graph mining together with visualization resulting in effective visual analytics for relational data. However, the visualization process is not a linear one as might suggest the plain reading of the mantras and pipeline. The analyst exercises its exploration cyclically iterating through Shneiderman and Keim's analysis/overview/zoom/details process. This is what makes visualization so different from graphical statistics and presents a real challenge. The back arrows in Fig. 1 actually encapsulate a complex process through which the user gains insight and understanding on the visualized data. A more user-centred depiction of the same visualization process is given in the NVAC document:

More recently, van Wijk suggested how to measure the effectiveness and benefits of a visualization in terms of learning efforts and acquired knowledge [96].

3.3. Visual Analytics for Graphs

Graphs offer a powerful and flexible mathematical tool to model real life phenomena. Biologists naturally use graphs to infer relationships between subcellular components (proteins, peptides, genes, RNAs, molecules, ...). Geographers have long used graphs to represent exchange networks (roads, air traffic, immigration,). Sociologists heavily rely on the use of graphs to study social networks. In all cases, the visual inspection of a network supports the analysis of its community structure and helps to answer questions concerning prominent actors (proteins; cities; manager; logical entity) or subgroups (biological function; territory; team; logical unit). The identification of communities in a network is an essential step towards understanding the whole network architecture. Once a subgroup has been identified, and when it appears as such within the visualization, it can be zoomed in to allow a more detailed inspection of its own dynamics. Graphs also appear as a natural modelling tool in computer science itself (data structures, web graphs, workflows, etc.).

Graphs moreover become a profitable metaphor when studying data equipped with a similarity measure either inherited from the data or computed from semantic attributes. A graph can readily be constructed applying a threshold on similarities. The use of a correlation measure to infer similarities is a common approach bringing similarities into the picture when analyzing data.

The case of image classification/indexing is typical. Structural indicators such as the MPEG-7 colour structure descriptors [78], [87] can be computed for each image; as a consequence, the similarity between any two images can be computed as a [0, 1] value. Images being highly similar can then be considered as neighbours in a (weighted) graph, enabling the analyst to exploit analytical tools borrowed from graph drawing, graph algorithmics, graph theory and combinatorial mathematics.

Bio-informatics also provide other interesting examples. For instance, an important use of DNA microarray data is to annotate genes by clustering them on the basis of their gene expression profiles across several microarrays. Because the transcriptional response of cells to changing conditions involves the coordinated co-expression of genes encoding interacting proteins, studying co-expression patterns can provide insights into the underlying cellular processes. In this context, the (Pearson) correlation coefficient is a standard dissimilarity measure used to infer network structure. On the assumption that genes and their protein products carry out cellular processes in the context of functional modules, it is natural to ask whether such modular organization can be revealed through the study of gene or protein interaction networks.

Graph Visualization is an active subfield of Information Visualization dealing with graph algorithms to find patterns, test properties, embed graphs in particular geometries (most often 2D or 3D Euclidean) or interactively manipulate their representations on the screen. Each year, a number of papers accepted at the IEEE InfoVis Symposium[3], the IEEE/Eurographics EuroVis Conference[4] or the IEEE Information

[2] See the Event Summary of the Workshop on Visual Analytics held at Konstanz University in June 2005: http://infovis.uni-konstanz.de/index.php?region=events&event=VisAnalyticsWs05

[3] See the URL http://www.infovis.org

Visualization Conference[5] concern graph visualization. The Graph Drawing community, with its own annual international symposium also contributes to the development of the field[6].

When focusing on relational data (graphs), combinatorial mathematics offer tools to exploit the topology of graphs and other structural regularities either numerically or from an algorithmic standpoint. A typical graph drawing algorithm will assume or test specific topological conditions such as being a tree or being bi-connected. Visualization techniques can benefit from combinatorial knowledge on particular graphs. One good example picked from our own results is the use of Strahler numbers (generalized to general directed or undirected graphs) to optimize the rendering of large graphs on a screen [41]. Other examples from our group exploit the fact that combinatorial parameters in a tree can be approximated using a Gaussian distribution [70], [44], folding or unfolding subtrees as the user navigates. Community identification methods based on using a node or edge dissimilarity measure in conjunction with a clustering method have proved fruitful.

The development and full exploitation of combinatorics to feed all subprocesses of the visualization pipeline (Fig. 1) with emphasis on the data analysis part is at the heart of our project. The core strength of our team resides in the development of combinatorial mathematics and graph algorithmics to serve the aims of graph visualization. We deploy our mathematical and algorithmic skills in Information Visualization to develop:

- Graph statistics: that capture key properties of the data, including scalable implementations;
- Clustering methods: that handle large datasets both visually and computationally;
- Graph hierarchies: that transform large graphs into a hierarchy of smaller, more readable and easier-to-manipulate sub-structures;
- Graph drawing algorithms: that lay out large datasets rapidly, enhancing scalability and addressing domain-specific conventions and requirements;
- Interactions: that exploit graph hierarchies as a central mechanism for navigating large graphs, while taking domain-specific tasks into account;
- Evaluation methods: that generate artificial datasets (randomly) based on key properties of the target data.

4. Application Domains

4.1. Introduction

> "**Collaborating with Application Domains**. To achieve greater penetration of visualization into application domains we must better integrate visualization capabilities with the requirements and environments of these domains. To achieve this integration, we must allow application goals, domain knowledge, and domain-specific conventions and metaphors to shape visualization methods. Visualization methods must address the characteristics of real, rather than ideal, data, addressing among others the challenges of heterogeneity, change over time, error and uncertainty, very large scale, and data provenance." [36]

Visualization is by nature fed by application domains. Questions are raised by and embodied in domain specific datasets. Working close to our end users is mandatory the users we are concerned with are experts from specific scientific domains or industrial sectors. Following an iterative process, the analysis/hypotheses evolve and refine while the visualization becomes more focused and adapted to the user's task. Put differently, the visualization helps expert users refine their hypotheses on the data while at the same time they become able to express their needs for visualization more effectively.

[4] See the URL http://www.eurovis.org
[5] See the URL http://www.graphicslink.co.uk
[6] See the URL http://www.graphdrawing.org

4.2. Social sciences

> "Visualization has thus far had less impact on the social sciences than the physical sciences, in part because of a dearth of funding for such efforts, but it holds the promise of effecting similar transformations." [36]

From our experience, we see social sciences as a fertile area from which ideas can emerge, and where visual analytics techniques and methods can be designed, tested and validated. Because social sciences deal with non deterministic phenomenon, it places us right in front of challenges underlined by Thomas and Cook: to deal with large volume and inhomogeneous data, with constant changes in data, possibly making it ambiguous and uncertain. It is part of our agenda to develop closer relationships with research teams or industrial partners in social sciences.

We have had the opportunity to work in close collaboration with experts from quantitative geography through the ANR SPANGEO (Masses de données 2005 call) project[7]. This ANR project has established close ties between individuals and long-term collaborations with the community research in geography. It gave us the opportunity to enter social sciences and explore the potentialities of interactive graph visualization and graph hierarchies for geographers [39], [54], [75], [52], [85], [84], [57], [76]. Our approach clearly appears as complementary to classical cartography.

Cartographers and geographers, because they often stick to the usual geographical world map to depict statistical data, are limited by the size of the dataset they can visualize and thus visually analyze. Graph visualization offers them the possibility of visualizing and navigating whole datasets, at the price of leaving aside geographical constraints.

Moreover, quantitative geography also offered us the occasion to compare graph combinatorics with tools and approaches based on graph theory developed by geographers. The theory of small world network as initiated by Watts and Strogatz [91], [93], [92] draws new insights on spatial analysis as well as to systems theory. Its concepts and methods are particularly relevant to geography where spatial interactions are mainstream, and where interactions can be described and studied using large volume of exchanges or similarities matrix. In terms of geographical analysis of spatial networks, our methodology helped expert identify network entities acting as bridges between several components and offer a higher capacity for urban communities to benefit from opportunities and create future synergies.

Multiscale models. Our methodology exploiting hierarchical graphs [43], [79], [6] appears as a fruitful strategy to discover scales in datasets [39].

Identifying structural changes. The data we study with quantitative geographers typically is time-stamped. That is, we often have data on populations, companies, air traffic, etc., collected through public surveys or by private companies over several years or months. The issues we now address is to identify structural changes or evolving patterns in networks. The task here is of great interest: the answer does not solely rely on mathematics or algorithms, but requires that experts link the identified pattern to real-life phenomenon and assess of its existence based on factual arguments (territorial policies, partnerships between companies, etc.).

Other partnership have been established with partners aiming at the development of visual approaches to hypothesis building and validation for law experts. The heterogeneous nature of the data will once again challenge us since law experts usually build cases based on information from texts (mail exchanges, newspaper extracts, etc.) and informal information from interviews or personal diaries.

4.3. Biology and bio-informatics

Because post-genomic data is made available publicly and is easily accessible on the web, biological data often becomes a target for visualization techniques we develop. Protein interaction networks nowadays are easy to find on the web; authors publishing in bio-informatics journals even sometimes provide the datasets they used as a benchmark.

[7] SPANGEO is a working group part of the S4 European initiative, see the URL http://s4.parisgeo.cnrs.fr/spangeo/spangeo.htm

Biology offers a fertile area for research in visualization because of massive data produced from experimentation, and also from a strong demand on the side of biologists. Problems in biology and bio-informatics concern almost all issues cited above:

- graph statistics enter the scene when inferring graph structures from biological data, or when interpreting measures/attributes associated with the data;
- the nested structure of metabolism naturally makes use of graph hierarchies.

Strong graphical conventions in biology (when drawing metabolic pathways or RNA secondary structures for instance) turn the design of automatic graph drawing algorithms into a real challenge. Part of our research was developed through the ANR AReNa project[8]. RNA comparisons required the development of new graph drawing algorithms and graph visualization software [64], [4]. The Bordeaux environment places us close to biologists and bio-informatics researchers. We also have contributed to the study of metabolism, more recently together with researchers from the INRIA HELIX project in Lyon [5]. The questions addressed through this starting collaboration this time concern the identification of biological patterns in metabolic or signal networks.

We have direct collaboration with biologists and bio-informatics researchers since the LaBRI team from which our project stems is directly concerned with bio-informatics, placing us close to the actual users (as with SPANGEO). Team members interact with researchers of the INRIA MAGNOME project; these interactions already gave rise to the development of the software application ProViz [72].

The ANR project SysTryp already gave us the occasion to reinforce collaborations with researchers from INRA Toulouse (Fabien Jourdan) and Bordeaux (Antoine de Daruvar, Patricia Thébault from CBib). SysTryp focuses on the relationship between metabolism and cellular differentiation in the protozoan Trypanosoma brucei, by collecting high resolution mass spectrometry data and reconstructing networks based on this data (see section 8.1).

4.4. Semantic networks and knowledge representation

Actors from the industry are now forced to store and filter huge amounts of information in digital formats on subjects concerning their domain of activity. This state of affairs calls for the development of technologies capable of helping human users to explore, filter and organize information and moreover discover concepts from non-structured or semi-structured data. After we had recognized here a situation where graph visualization could help, we grouped with industrial actors in a technology transfer effort. Although the industry can certainly benefit from our actual expertise, the project nevertheless requires genuine research.

Our research in this field is conducted in close collaboration with the industry and focuses on developing techniques and tools devoted to competitive and strategic watch. After a successful experience working with PIKKO and AMI Software[9] through a 2 year ANR RNTL project, we recently grouped our efforts with Thalès Communications and Xerox. The TANGUY[10] project is funded through the ANR CONTINT call. This project is concerned with processing legal text documents, mixed with unstructured and uncertain information gathered from various sources into a semantic network. The project focuses on developing a genuine approach to help legal experts to explore dense legal documents.

Our efforts focus on building and delivering answers from visual queries on a semantic network. We also plan to provide tools to visually support reporting activities. The project requires the design of astute graphical representations and data handling.

We see in this project a real opportunity to address issues raised within the knowledge representation domain. In a sense, this project should reinforce our expertise on the processing and visualization of dynamic data.

4.5. Theoretical Computer Science – Formal methods / Graph rewriting systems

Participants: Jonathan Dubois, Hélène Kirchner, Guy Melançon, Bruno Pinaud.

[8] See the URL http://www.lri.fr/~denise/AReNa/
[9] See the URLs http://www.pikko-software.com and http://www.amisw.com.
[10] TANGUY stands for "From Text to Arguments through Networks with Goals and User Initiative".

Designing visualization techniques and tools to explore dynamic graphs holds a central place in our research agenda. Since early 2009, we had the opportunity to start a collaboration with experts on graph rewriting systems.

That is, the challenge here is to be able to visually support the exploration and analysis of graphs whose topologies evolve in time according to rewriting rules. Our goal ultimately is to develop a system allowing to visualize the graphs as they evolve, and also offer a complete environment to study the rewriting system in itself. In such a setting, the visualization supports the study of the rewriting system, helping the identification of properties such as its convergence or non-convergence, deadlock situations, frequent or rare instances, etc. On top of all visualization issues it raises, the system also looks at pattern matching and computing history issues. We also look at issues related to the internal representation of the system required for its visualization.

5. Software

5.1. Tulip: a scalable Graph Visualization Framework

Participants: David Auber [correspondant], Jonathan Dubois, Ludwig Fiolka, Morgan Mathiaut, Patrick Mary.

See also the web site at http://www.tulip-software.org.

Most of our work requires experimentation and validation. To this end, our group uses the Tulip graph visualization framework designed and developed by David Auber [42]. Tulip offers a C++ plug-in mechanism easing the development and adjunction of new algorithms (computing graph statistics, graph drawing or graph clustering). This actually is a main feature of the Tulip platform [46]. We do however experiment with other information visualization platforms such as the InfoVis Toolkit developed at INRIA [62] and prefuse [68] both based on Java technology.

Tulip implements astute data structures, data management and filtering/inheritance mechanisms. Its rendering engine relies on OpenGL, while its GUI rests on Trolltech's QT library. The main interaction paradigm offered by Tulip is the computation and direct manipulation of graph hierarchies, making it unique among all available graph visualization platforms such as Pajek [51], [94], [65], [60], [61] and others.

Tulip's architecture eases the construction of stand-alone and specialized applications. An application targeted at domain specific data can be realized through the selection of building blocks (data structures, existing plug-ins), customization/development of a QT-based GUI, and adjunction of specific plug-ins. Tulip's popularity is confirmed by its (approximately) 1000 downloads / month since it was ported on sourceforge. Tulip is also shipped with several Linux releases. Its unique C++/Linux plug-in mechanism has recently been the focus of a short article in a wide audience French magazine [46].

We can list several applications that were built from Tulip:

- EVAT was explicitly designed for the navigation and comparisons of huge trees (file systems, classification of species) and won our team a second place at the InfoVis Annual Contest [45];
- SWViz implemented a specific hierarchical clustering algorithm for multilevel visualization and navigation of small world networks (social networks, software re-engineering graphs, etc.) [43];
- ARNa implements an astute planar graph drawing algorithm capable of producing drawing of RNA secondary (2D) structures that mimics hand made drawing by biologists; the software is moreover designed to support the visual comparison of two RNA structures, finding the most common substructures and deciding of a best viewpoint on the drawings [64];
- Tulip again was second place at the InfoVis Annual Contest in 2004, helping to explore and understand the organization and evolution of an information space made from InfoVis publications, author, topics,...over the past ten years [59].

We should also mention the use of Tulip framework without declining it into a specific application by other research teams. The Sorenson Molecular Genealogy Foundation studies population genealogy from a molecular biology perspective [81], where graph clustering is astutely used to trace back genealogy from DNA data. Roman *et al* [83] use Tulip to support their work on many-assets market modelling. Because of its plug-in mechanism, and of its openness, Tulip has been adopted by quite a large part of the Linux community - it is actually distributed in several Linux releases. Lately, Tulip was used to visualize and analyze a medieval social network [53] (see also http://blogs.zdnet.com/emergingtech/?p=928).

Additionally, although Tulip was originally intended to be used by a single user on a desktop workstation, its modular architecture allows us to explore its use in a collaborative setting [49], [48], [47].

5.1.1. Perspectives

Our team develops and uses Tulip for its own needs. Tulip is more than just a visualisation application, although the framework is shipped with an already usable user interface. Tulip has received much attention from the scientific community [40], [42].

Software modules from Tulip can be used separately to build customized applications requiring acute graph data management or basic graph visualization algorithms. Based on C++, OpenGL and QT from Trolltech, it makes use of the full capabilities of standard desktop computers now equipped with high-end graphics hardware.

We already have listed applications that were built from Tulip. Our work with geographers within the ANR project SPANGEO should lead to the development of a specific application dedicated to the visualization and navigation of spatial graphs in quantitative geography. This should allow us to enter the scene in the field of social network analysis, competing with Pajek [94], for instance.

Part of Tulip's evolution will certainly rely on our ability to feed the framework with new algorithms and visualization techniques, as it has been the case since now. But there is a more challenging evolution we wish to address.

At the moment, Tulip appears as a rich collection of algorithms easing experimentation with large graphs. The next step is the construction of a development kit "à la Eclipse"[11], enabling users to switch between perspectives, going from a pure algorithmic perspective to a task-centered one. A perspective here must be understood as a collection of GUI tools and associated plug-ins just as what can be found for Eclipse.

Selecting a "file system" perspective could bring together specific graph drawing algorithms and graph statistics, graph clustering algorithms, relevant interactors, etc. Other perspectives are possible: social network analysis, RNA comparisons, software re-engineering or software visualization, etc. We are confident that our experience in designing domain specific application from Tulip will help.

This subproject is a real challenge in itself in that it will most certainly require the adoption, development and/or adjustment of a taxonomy for data, tasks and/or application domains from which perspectives "à la Eclipse" can be defined [77], [58], [74], [86], [56], [50], [55], [69], [89], [63], [88], [73].

It will most certainly require the development of a more ambitious type of plug-in, that we might call "domain specific" plug-in, or "task-oriented" plug-in, meaning that it combines proper algorithmics, drawing and interaction relevant for a given application domain or given type of tasks offering, a kind of bouquet of more elementary plug-ins combining into a *visualization environment*.

The actual architecture of Tulip, assembling independent modules, seems ready for this type of evolution without requiring a total rewrite of the framework. A web service project (Tulip plugin manager web service[12]) has already been created on sourceforge with that mid and long-term evolution of Tulip in mind.

[11] Eclipse is an open source SDK primarily developed and maintained by IBM, originally intended for the Java developer community. Its success basically emerges from its ability of integrating various plug-ins, offering services such as name completion and easy editing, code versioning, debug, etc.

[12] See the URL http://sourceforge.net/projects/tulip-plugin-ws/

We also should underline the impact and benefit of such an approach for the industry, easing the adoption or integration of Tulip's technology into their own. The development of perspectives could also evolve into commercial products some Eclipse perspectives already are.

5.2. GVSR: Graph Visualization Software References

Participant: Bruno Pinaud [correspondant].

See http://gvsr.polytech.univ-nantes.fr.

The profusion of available graph visualization applications may even confuse an expert in this field. Some programs have been developed in close partnership with the scientific community (Pajek [94], Cytoscape[13]), others are purely commercial, or some are general graph manipulation and visualization software (Tulip). Generally speaking, the choice of a program well-adapted to both the data and the methodology is difficult. Some books can be used as guides [37], [80], and several websites present lists of programs (e.g. http://rw4.cs.uni-sb.de/users/sander/html/gstools.html, http://www.manageability.org/blog/stuff/open-source-graph-network-visualization-in-java/). However, those websites plainly list the existing software, or make them accessible through snapshots. Consequently much effort is required to compare the various programs before choosing the best one for the problem considered.

Those observations led us to develop GVSR (http://gvsr.polytech.univ-nantes.fr). Its added value is to offer users query about existing software based on commonly used criteria such as scalability, implementation issues or type of uses. Our objectives are to facilitate the users' choices and to compare programs with common criteria. The website also presents the programs with a uniform text-based description. This site keeps evolving and so far contains eighty various software descriptions. In addition, the site allows users to propose new programs by simply completing an enclosed form. The site is also designed as a tool repository helping the user to access and compare the available Graph Visualization tools, and benchmark new techniques and algorithms. The whole community can benefit from the ability to reproduce published results, and from comprehensive comparisons with previous work. Thus, GVSR can be seen as a contribution to improving both the accessibility and quality of graph visualization tools.

6. New Results

6.1. Graph Drawing

Participants: David Auber, Romain Bourqui, Antoine Lambert, Arnaud Sallaberry, Paolo Simonetto.

6.1.1. Edge Bundling

Visualizing graphs containing many nodes and edges efficiently is quite challenging. Drawings of such graphs generally suffer from visual clutter induced by the large amount of edges and their crossings. Consequently, it is difficult to read the relationships between nodes and the high-level edge patterns that may exist in standard node- link diagram representations. Edge bundling techniques have been proposed to help solve this issue, which rely on high quality edge rerouting. We introduce in [16] an intuitive edge bundling technique which efficiently reduces edge clutter in graphs drawings. Our method is based on the use of a grid built using the original graph to compute the edge rerouting. In comparison with previously proposed edge bundling methods, our technique improves both the level of clutter reduction and the computation performance. The second contribution of this paper is a GPU-based rendering method which helps users perceive bundles densities while preserving edge color.

In [24], we present a generalization of [16] to reduce the clutter in a 3D representation by routing edges into bundles as well as a GPU-based rendering method to emphasize bundles densities while preserving edge color. To visualize geographical networks in the context of the globe, we also provide a new technique allowing to bundle edges around and not across it.

[13] see http://www.cytoscape.org

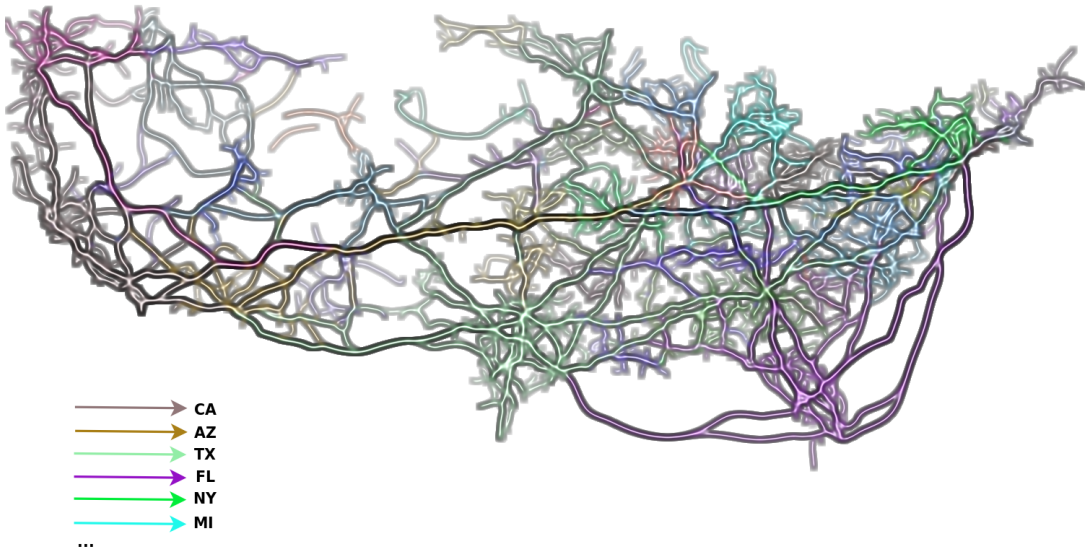

Figure 2. US migration graph visualization with our technique, heights are linearly mapped to the splat field and the diffuse map used for the bump mapping rendering corresponds to the splat field linear color mapping.

Figure 3. 3D World Air Traffic visualization with our technique.

6.1.2. Pattern visualization

Data mining techniques allow users to discover novelty in huge amounts of data. Frequent pattern methods have proved to be efficient, but the extracted patterns are often too numerous and thus difficult to analyze by end-users. In [26], we focus on sequential pattern mining and propose a new visualization system, which aims at helping end-users to analyze extracted knowledge and to highlight the novelty according to referenced biological document databases. Our system is based on two visualization techniques: Clouds and solar systems. We show that these techniques are very helpful for identifying associations and hierarchical relationships between patterns among related documents. Sequential patterns extracted from gene data using our system were successfully evaluated by two biology laboratories working on Alzheimer disease and cancer.

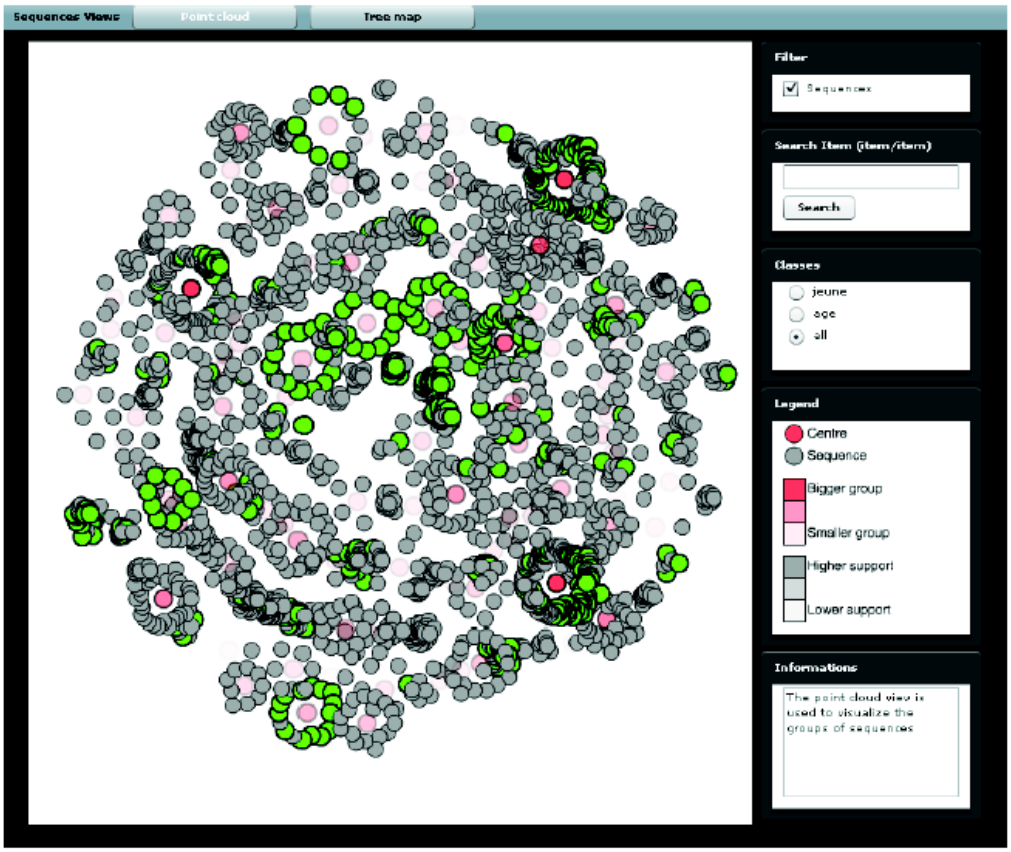

Figure 4. Point cloud with sequences and highlighted researched elements.

6.1.3. Hypergraph

A path-based support of a hypergraph H is a graph with the same vertex set as H in which each hyperedge induces a Hamiltonian subgraph. While it is NP-complete to compute a path-based support with the minimum number of edges or to decide whether there is a planar path-based support, we show in [20] that a path-based tree support can be computed in polynomial time if it exists.

In [19], we show how to test in polynomial time whether a given hypergraph has a cactus support, i.e. a support that is a tree of edges and cycles. While it is NP-complete to decide whether a hypergraph has a 2-outerplanar

support, we show how to test in polynomial time whether a hypergraph that is closed under intersections and differences has an outerplanar or a planar support. In all cases our algorithms yield a construction of the required support if it exists. The algorithms are based on a new definition of biconnected components in hypergraphs.

6.2. Graph statistics and graph clustering

Participants: Daniel Archambault, Romain Bourqui, Maylis Delest, Frédéric Gilbert, Guy Melançon, François Queyroi, Arnaud Sallaberry, Paolo Simonetto, Faraz Zaidi.

6.2.1. Community detection in static networks

Searching of information on the web is a frequent task requiring some sort of organization to facilitate the searching process. Often this information is distributed, semistructured, overlapping and heterogeneous. Organization and structuring this information is an active area of research where the goal is to help users locate required information efficiently. Clustering is a well known technique to group similar information. Although often described as the unsupervised learning, clustering is quite trivial, as it often requires human intervention in terms of a number of parameters to guide the process. We address the clustering problem of web pages where the goal is to organize information to facilitate users for faster access to required information. In [30] We introduce a hierarchical fuzzy clustering algorithm to organize web pages. The algorithm uses a topological decomposition on the co-occurrence network of keywords to devise heuristics which help determine the input parameters for our clustering algorithm. Finally, we compare the results of the proposed algorithm with existing algorithms in the literature.

The exponential growth of data in various fields such as Social Networks and Internet has stimulated lots of activity in the field of network analysis and data mining. Identifying Communities remains a fundamental technique to explore and organize these networks. Few metrics are widely used to discover the presence of communities in a network. We argue that these metrics do not truly reflect the presence of communities by presenting counter examples. This is because these metrics concentrate on local cohesiveness among nodes where the goal is to judge whether two nodes belong to the same community or vise versa. Thus loosing the overall perspective of the presence of communities in the entire network. In [29], we propose a new metric to identify the presence of communities in real world networks. This metric is based on the topological decomposition of networks taking into account two important ingredients of real world networks, the degree distribution and the density of nodes. We show the effectiveness of the proposed metric by testing it on various real world data sets.

6.2.2. Community detection in dynamic social networks

Detection of community structures in social networks has attracted lots of attention in the domain of sociology and behavioral sciences. Social networks also exhibit dynamic nature as these networks change continuously with the passage of time. Social networks might also present a hierarchical structure led by individuals who play important roles in a society such as managers and decision makers. Detection and visualization of these networks that are changing over time is a challenging problem where communities change as a function of events taking place in the society and the role people play in it. In [15], we address these issues by presenting a system to analyze dynamic social networks (see Fig. 7). The proposed system is based on dynamic graph discretization and graph clustering. The system allows detection of major structural changes taking place in social communities over time and reveals hierarchies by identifying influential people in social networks. We use two different data sets for the empirical evaluation and observe that our system helps to discover interesting facts about the social and hierarchical structures present in these social networks.

In [35] we give the complete description of the graph decomposition algorithm used in [15] to generate overlapping clusters. The complexity of this algorithm is $O(|E|{\cdot}deg_max^2 + |V|{\cdot}log(|V|)))$. This algorithm is particularly efficient due to its ability to detect major modifications along dynamic processes such as time related ones.

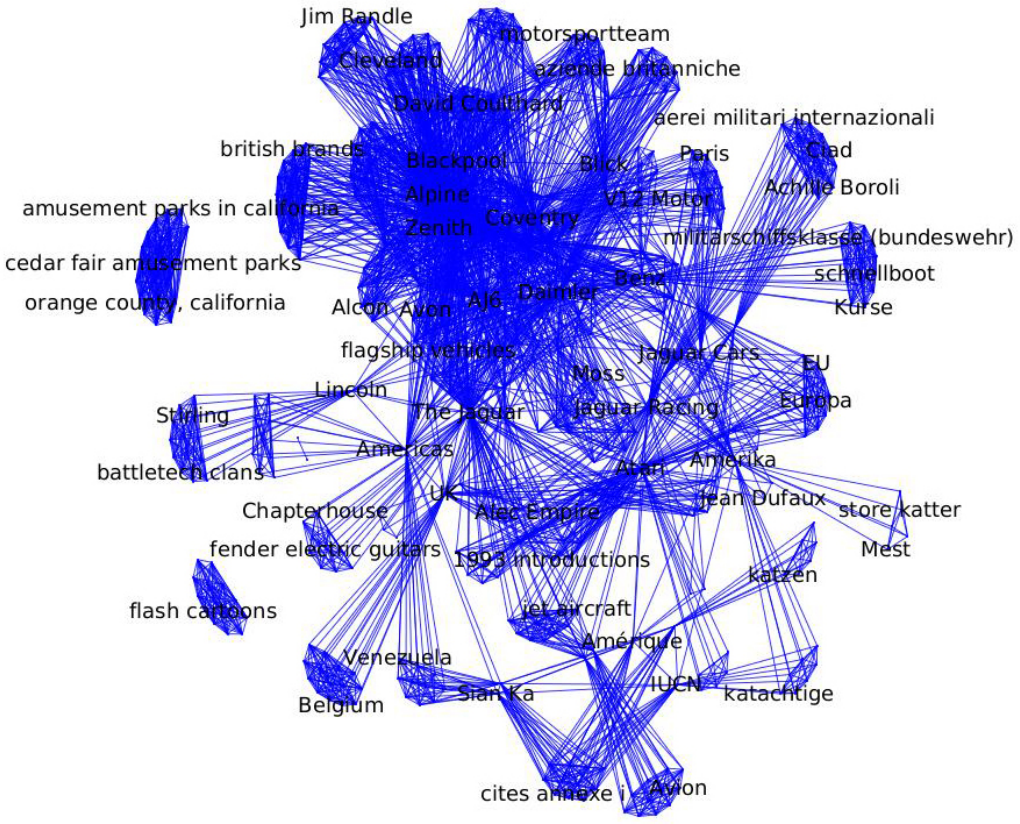

Figure 5. Co-occurrence network for Jaguar keyword where connected components can be easily identify as forming a clique.

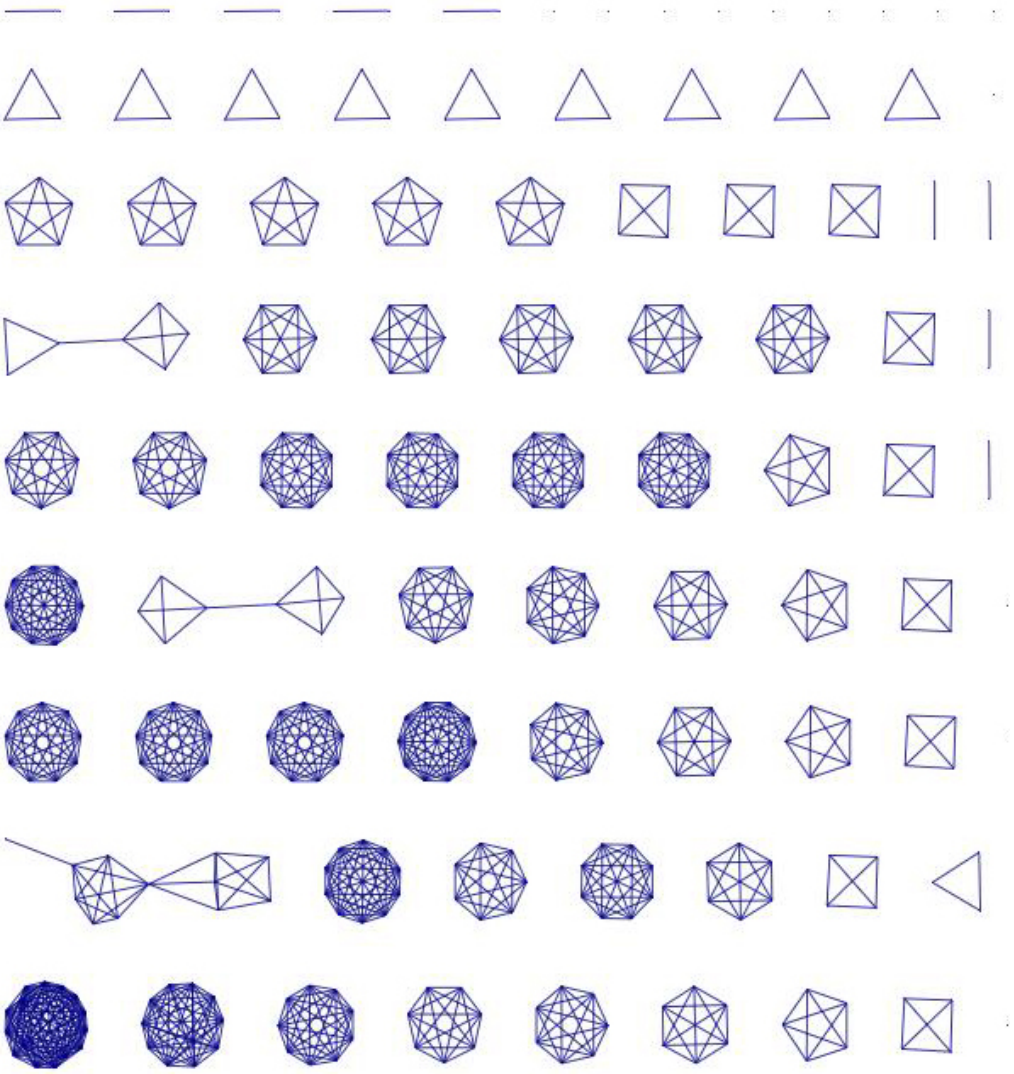

Figure 6. The Max16-DIS for the IMDB graph representing lots of densely connected components.

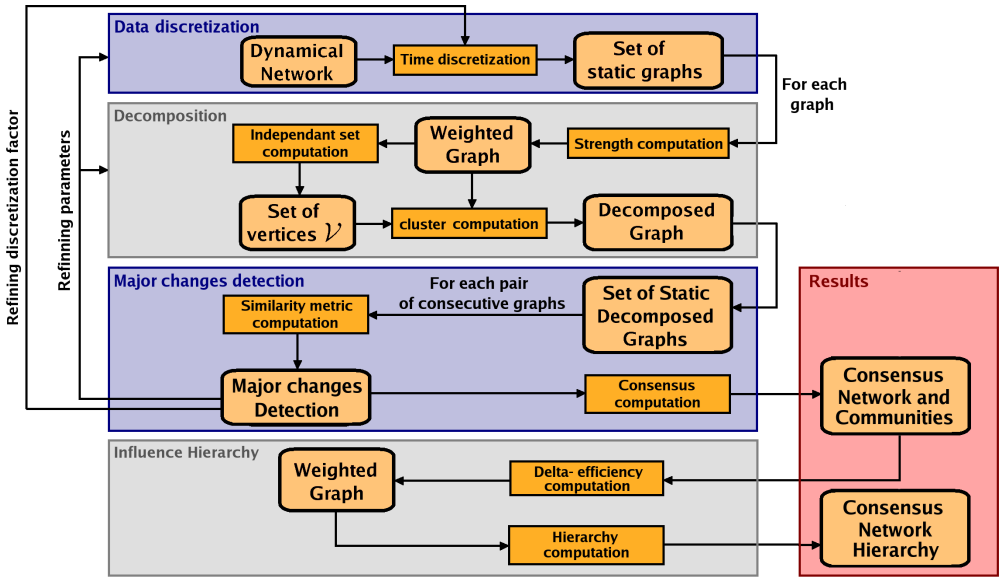

Figure 7. Community detection in social networks – Framework overview

6.2.3. Evaluation of clustering quality

Many real world systems can be modeled as networks or graphs. Clustering algorithms that help us to organize and understand these networks are usually referred to as, graph based clustering algorithms. Many algorithms exist in the literature for clustering network data. Evaluating the quality of these clustering algorithms is an important task addressed by different researchers. An important ingredient of evaluating these clustering techniques is the node-edge density of a cluster. We argue that evaluation methods based on density are heavily biased to networks having dense components, such as social networks, but are not well suited for data sets with other network topologies where the nodes are not densely connected. Example of such data sets are the transportation and Internet networks. We justify our hypothesis by presenting examples from real world data sets.

In [28], we present a new metric to evaluate the quality of a clustering algorithm to overcome the limitations of existing cluster evaluation techniques. This new metric is based on the path length of the elements of a cluster and avoids judging the quality based on cluster density. We show the effectiveness of the proposed metric by comparing its results with other existing evaluation methods on artificially generated and real world data sets.

In [33], We design and study a multilevel modularity quality for clustered graphs, explicitly taking the nesting structure of clusters into account. Multilevel models appear crucial in the natural and social sciences. The multilevel modularity quality measure generalizes a modularity quality measure introduced by Mancoridis in the context of reverse software engineering. The measure we designed recursively traverses the hierarchy of clusters and computes a one variable polynomial encoding the intra and inter-cluster connectivity ratios appearing at all levels in a hierarchical clustering. The resulting polynomial reflects how the graph combines with the hierarchy of clusters and can be used to assess the quality of a hierarchical clustering. We discuss examples as proof-of-concept.

6.3. Visual Analytics

Participants: Daniel Archambault, Romain Bourqui, Frédéric Gilbert, Guy Melançon, Paolo Simonetto, Faraz Zaidi.

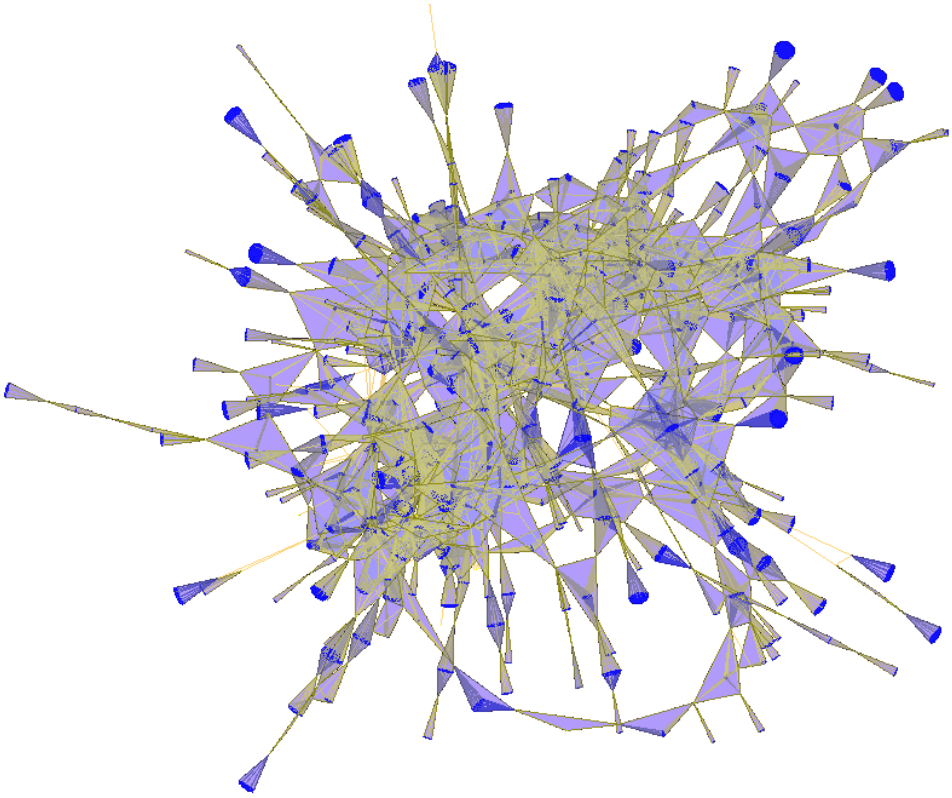

Figure 8. Result of our decomposition algorithm on a subgraph of the "Hollywood graph" (actors graph) containing 421 movies. Our algorithm detected 404 of these movies.

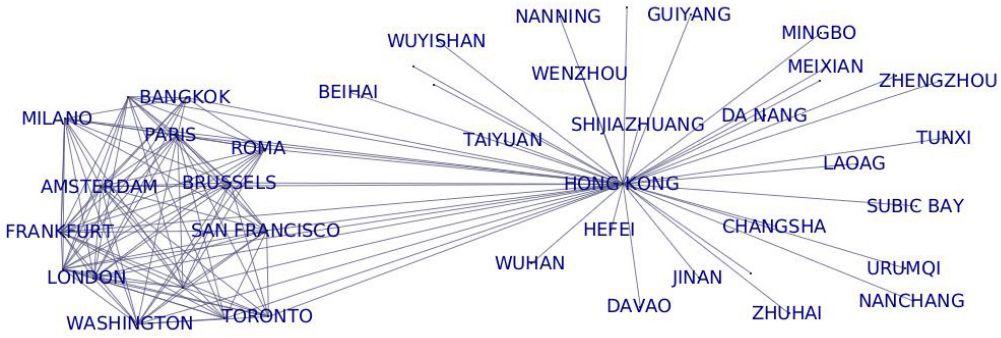

Figure 9. Air Traffic Network Drawn using Hong Kong at the center and some airports directly connected to it.

6.3.1. Interactive graph mining

Interactive graph mining is what we aim at designing and realize. Efficient visual analytics requires to astutely combine interaction with together with graph statistics and graph drawing. Building effective visualization systems is difficult as it requires to combine analytics based on data analysis brought to understandable and intuitive graphical representations equipped with adequate user interaction.

The use of graph hierarchies still remains a central paradigm we exploit. Often after graph hierarchy construction, there exists several large metanodes which can contain tens or hundreds of thousands of nodes. In steerable graph drawing systems, these metanodes can be a problem as they could take a half an hour or more to draw in their entirety. However, frequently, a user is interested in nodes that are close to a particular node or subgraph present in the hierarchy.

We have improved TugGraph, a system for exploring paths and proximity around nodes and subgraphs part of a hierarchy of subgraphs [11]. The approach modifies a pre-existing hierarchy in order to see how a node or subgraph of interest extends out into the larger graph. The system works well on graphs of hundreds of thousands of nodes and millions of edges and is able to present this information in a matter of seconds. TugGraph is a follow-up of previous work published in 2008 [2].

6.4. User Evaluation

Participants: Daniel Archambault, Bruno Pinaud.

The ACM special interest group on Computer-Human Interaction (HCI) "ACM SIGCHI Curricula for Human-Computer Interaction" proposes a definition of HCI (see http://old.sigchi.org/cdg/):

> *Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.*

We already work in the design and implementation of interactive techniques for visualization and interaction with graphs. More generally, many techniques and systems have been proposed in the InfoVis community but there is a lack of user evaluation of those systems. Thanks to close collaboration with Daniel Archambault (lique Strategic Research Cluster at UCD Dublin) and Helen Purchase (Department of Computing Science, University of Glasgow) we did three differents user evaluation. The data sets used in the experiments were derived from standard benchmark data sets of the information visualization community. The Questions used were selected to test both local and global properties of the graphs.

6.4.1. Animation, Small Multiples and the Effect of Mental Map Preservation in Dynamic Graphs

The majority of dynamic graph drawing algorithms have been designed with an animation of the sequence of graphs as their output in mind. In the information visualization community, small multiples has been in use for many years to display dynamically evolving data. In a small multiples approach, a matrix of images shows the differences between objects. In the case of dynamic data, each image in the matrix is a timeslice. Users are able to see and compare all timeslices at the same time.

We compared the performance of the animation of dynamic graphs to the presentation of small multiples and the effect that mental map preservation had on the two conditions. We found that small multiples gave significantly faster performance than animation overall and for each of our five graph comprehension tasks. In addition, small multiples had significantly more errors than animation for the tasks of determining sets of nodes or edges added to the graph during the same timeslice, although a positive time-error correlation coefficient suggests that, in this case, faster responses did not lead to more errors. This result suggests that, for these two tasks, animation is preferable if accuracy is more important than speed. Preserving the mental map under either the animation or the small multiples condition had little influence in terms of error rate and response time.

More details can be found in [12].

6.4.2. The readability of Path-Preserving Clusterings of Graphs

Graph visualization systems often exploit opaque metanodes to reduce visual clutter and improve the readability of large graphs. This filtering can be done in a path-preserving way based on attribute values associated with the nodes of the graph. Despite the extensive use of these representations, as far as we know, no formal experimentation exists to evaluate if they improve the readability of graphs. We ran a user study that formally evaluates how such representations affect the readability of graphs. We also explore the effect of graph size and connectivity in terms of this primary research question. Overall, for our tasks, we did not find a significant difference when this clustering is used. However, if the graph is highly connected, these clusterings can improve performance. Also, if the graph is large enough and can be simplified into a few metanodes, benefits in performance on global tasks are realized. Under these same conditions, however, performance of local attribute tasks may be reduced.

More details can be found in [18].

6.4.3. Difference Map Readability for Dynamic Graphs

Difference maps are one way to show changes between timeslices in a dynamic graph. They highlight, using colour, the nodes and edges that were added, removed, or persisted between every pair of adjacent timeslices. Although some work has used difference maps for visualization, no user study has been performed to gauge their performance. We did a user study to evaluate the effectiveness of difference maps in comparison with presenting the evolution of the dynamic graph over time on three interfaces. We found evidence that difference maps produced significantly fewer errors when determining the number of edges inserted or removed from a graph as it evolves over time. Also, difference maps were significantly preferred on all tasks by our users.

More details can be found in [17].

7. Contracts and Grants with Industry

7.1. RAISME: Rapid Application Innovation for SMEs

Participants: David Auber, Morgan Mathiaut.

- Project: RAISME, Rapid Application Innovation for Services For Multidimensional Enterprises
- Call: call 3 FP7-SME-2010-1
- start/end november 2010 – October 2012
- Budget: 170 000 euros (grant for INRIA) euros (INRIA GRAVITÉ)

This project is an innovative collaboration between three high-tech SMEs with global aspirations working with arguably three of the best RTD performers in Europe. The project itself and the outcomes resulting from the work will have a unifying theme of innovation, flexibility, collaboration, and of delivering whole product solutions.

In technology markets the goal is market share via platform solutions. At the start of the development life cycle the whole product barely exists. At best there will be a core product surrounded by an envelope of custom services required to make the application work. In this early market visionaries commit to an incomplete product and use it as a foundation for fielding an application breakthrough. However, the challenge of driving out the service component and of providing a fully integrated commoditised product, where significant revenue can fund the building of valuable market share is beyond the reach of most SMEs. In fact innovation by SMEs is often characterized by products with a low R&D intensity and a limited view of markets, and very few companies ever establish products of international scale.

The RAISME project will enable high-tech SMEs with niche skills to rapidly build and scale innovative ICT applications through the collaborative use of advanced “mashup” technology and cloud computing. The RAISME SMEs will be at the vanguard of a new business paradigm where the end-user becomes part of the product development lifecycle, thus accelerating it, and where importing innovations from the internet allows faster exploitation of knowledge and productivity. The RTD performers who have already developed state-of-the-art visualisation, optimisation and integration tools will boot-strap a series of niche applications brought to the project by the SMEs to then deliver highly innovative knowledge services to the market. Additionally the anticipated project outcomes will help mobilise communities of high-performing SMEs to compete in worldwide markets.

7.2. Semantic networks and knowledge representation

Participants: David Auber, Ludwig Fiolka, Antoine Lambert, Frédéric Gilbert, Guy Melançon, Arnaud Sallaberry, Faraz Zaidi.

- Project: TANGUY From Text to Arguments through Networks with Goals and User Initiative
- Call: ANR CONTINT (Content and Interaction)
- start/end January 2009 – December 2011
- Budget: 872 000 euros (total) / 349 000 euros (grant) / 261 000 euros (INRIA GRAVITÉ)

We got involved in a project with two industrial partners, Thalès Communications[14] and the Xerox XRCE Parsing and Semantics group[15], and a third partner acting as “final user” – the FIDAL Law Firm.

This new ANR CONTINT project – TANGUY – aims at providing technological solutions to users confronted with ever growing amounts of information. Today the information overload is mainly handled by indexing data sources in search engines. In order to increase the relevance of the results, indexation is enriched by natural language processing (NLP) tools, especially by tools of information extraction. Another aid are clustering tools that give synthetic views of the results.

Recently, semantic web research has been using knowledge representation tools to organize information along the lines of ontologies or semantic networks. Some research has been carried out to integrate information extraction tools and knowledge representation tools. In practice, the knowledge representation approach leads to static, "engineering"-centered architecture, i.e. all the knowledge needed is fixed in advance by technical experts who are not connected to the final users. The TANGUY project tackles the problem in a quite different way. It intends to yield the users the most information and independence possible in processing their individual issues and at the same time it aims at reducing processing time and engineering costs. The TANGUY approach consists in the symmetrical and dynamic cooperation among three "poles":

1. Extraction of « micro knowledge elements » from the texts with the help of NLP tools. This extraction leads to structuring the texts into elementary pieces locally with respect to a particular document, and not specialized in a predefined domain.
2. Knowledge representation with the help of semantic networks. This pole constructs an overall vision of inter-documentary relationships in the corpus. It supports navigation, query and reasoning operators that lead the user towards his goal. It also allows the creation of a network of high-level concepts that represent the users’ understanding.
3. Interaction with the semantic network. This pole gives way to exploring at various levels the continuum between the information extracted and the concepts that synthesize them. This is the tool that guides the actions (incorporation of new texts, exploration and creation of new concepts) in order that the user can construct and argument the solution of his current problem.

[14] See the URL http://www.thalescomminc.com
[15] See the URL www.xrce.xerox.com/Research-Development/Document-Content-Laboratory/Parsing-Semantics

8. Other Grants and Activities

8.1. ANR Systryp

Participants: David Auber, Romain Bourqui, Jonathan Dubois, Paolo Simonetto.

- Project: SysTryp (Metabolomic systems biology analysis of differentiation in trypanosomes)
- Call: ANR Systems Biology (bilateral FR-UK)
- start/end December 2007 – December 2010
- Budget: 299 980 euros (grant French partners) / 89 338 euros (INRIA GRAVITÉ)

The project focuses on the study of the relationship between metabolism and cellular differentiation in the protozoan Trypanosoma brucei, by collecting high resolution mass spectrometry data and reconstructing networks based on this data. Relationships between static and dynamic networks will be determined and hypotheses generated by seeking and visualizing metabolic network modules that associate with differentiation.

A limitation in modeling of biochemical networks relates to a lack of general compatibility between static and dynamic modeling. Here we aim to reduce this gap and provide the means by which biochemists move seamlessly from the global view of metabolism within a model system, provided by static modeling, to a detailed representation derived from dynamic modeling. To do so, we will design and evaluate new combinatorial and visual means to detect, within large networks, modules corresponding to key pathways involved in the system under study. To validate these graph mining methods we will model one selected pathway using dynamic modeling and then check it experimentally. We will focus on the protozoan, Trypanosoma brucei, an extraordinary model system. These single celled organisms undergo a complex life cycle that takes them through the divergent environments of the mammalian bloodstream through various developmental stages within the tsetse fly. As a consequence the trypanosome remodels its cellular structure, and its metabolism, to adapt to these incongruent conditions.

Once within those environments, however, they enjoy relative stability, thus their capacity to retain homeostasis is apparently pre-programmed and their metabolic network less plastic than those seen in free living organisms like yeast. Here we propose to make comprehensive measurements of the trypanosome's metabolome as the parasites transform. Ab initio networks, where individual metabolites are linked based on chemical transformations between them, will be constructed along with a second set of networks of metabolites whose abundance changes in a coordinated fashion. The various networks will be used to assist in validating the accuracy of the overall network. Modules, comprising connected metabolites whose abundance changes in a coordinated fashion throughout the differentiation process will be identified and the components of a selected module will be subject to dynamic modeling. Predictions based on the modeling will then guide reverse genetics based experiments (using gene knockout and RNA interference) to remove genes encoding enzymes central to the modules predicted to be critical to differentiation. The impact of these genetic perturbations on the differentiation process and on the metabolome will be assessed experimentally.

In summary, the project aims:

1. To use high resolution mass spectrometry to identify the metabolite composition of trypanosomes and see how the metabolome changes during the differentiation process;
2. To use advanced bioinformatic techniques (based on metabolic connectivity and response correlation) to build metabolite networks from these cells;
3. To follow perturbations, and use static modeling to identify those parts of the network that are unchanged through the differentiation processes and those which respond, in a coordinated fashion, to changes during differentiation;
4. To use modularity properties to derive visualisation methods that allow identification of one or more modules associated with cellular differentiation, then use them in dynamic flux modeling studies to evaluate their cellular roles;
5. To use a functional genomics approach to find whether loss, or inappropriate expression of key enzymes involved in differentiation-response pathways have a predictable impact on this process.

8.2. "Equipe associée" PORGY

Participants: Jonathan Dubois, Hélène Kirchner, Guy Melançon, Bruno Pinaud.

- Project: PORGY (Interactive Analysis and Visualisation of Port Graph Rewriting Systems)
- Call: INRIA Associate Team call
- start/end January 2009 – December 2011
- Budget: 26 900 euros (grant for all partners)

PORGY aims at designing genuine techniques to support the visual exploration and analysis of graph rewriting systems. A graph rewrite system is a set of rewrite rules of the form $A \rightarrow B$ where A and B are graphs with variables. Such a rule applies to a graph G if G contains at least one instance of the left-hand side A, i.e. a subgraph homomorphic to A. Then G rewrites to a new graph G' obtained by replacing the instance of A by an instance of B where variables have been instantiated as in A. This induces a transitive relation on graphs. Each rule application is a rewriting step and a derivation is a sequence of rewriting steps, that we will sometimes call a computation, referring to application of rewriting to programming languages.

From a visualization perspective, dealing with graph rewriting systems boils down to dealing with dynamic graphs – although the dynamic aspects here are not given as time-stamped evolution of a graph but rather are consequences of the non-deterministic applications of rewriting rules. indeed, rewriting is intrinsically non-deterministic since it may be possible to rewrite several subgraphs of a graph at the same time with different rules or the same one, possibly getting different results.

Rewriting systems have been studied as theoretical apparatus in computer science, as a rewriting system can be considered as a "program", in a sense. These formal objects however have interesting potential applications.

A port graph is a graph where nodes have explicit connection points called *ports* and the edges are attached to specific ports of nodes. This type of graph was identified from an abstract view of proteins and the molecular complexes resulted from the protein-protein interactions in a biochemical network. From a biochemical perspective, a protein is characterized by a collection of small patches on its surface, called functional domains or sites, and one kind of protein-protein interaction results in the two proteins binding on some particular sites. Then a protein, viewed as a named collection of sites, is graphically modeled by a node with ports, and a bond between two proteins by an edge in a port graph. Graphically, a node in a port graph is represented as an empty box having the identifier placed at the exterior and the ports as small points on the surface of the box. The rewriting of subgraphs correspond here to a change of states of bio-molecules. The study of the rewriting systems becomes critical to gain a better understanding of the biochemical mechanism driving interactions between molecules.

A user of the Porgy system can describe with a simple programming language how the rewriting rules have to be applied. Thanks to Tulip, the user only needs to select a source model and write the strategy. Then, all the rules are triggered and the trace tree is updated accordingly. The strategy language includes operators to select the rules to apply and the positions where the rules are applied. Those positions can also be changed along the derivation.

The PORGY project gathers researchers from GRAVITÉ[16] and King's college London[17].

After designing a software architecture allowing the interactive manipulation and visualization of graph rewriting systems during the first year of the project, efforts have been devoted to the development of an editor easing the generation of all necessary components (see Fig. 10). All graphs produced along the application of rules are stored into a tree of graph instances thus relying on a node/metanode approach. Rules are themselves considered as graphs. The visual representations of all these components raise a number of graph drawing and graph visualization issues we are currently studying.

[16] The PORGY "Equipe associée" also includes Hélène Kirchner, although Hélène is not a permanent member of GRAVITÉ.
[17] See the URL http://gravite.labri.fr/?Projects_%2F_Grants:Porgy.

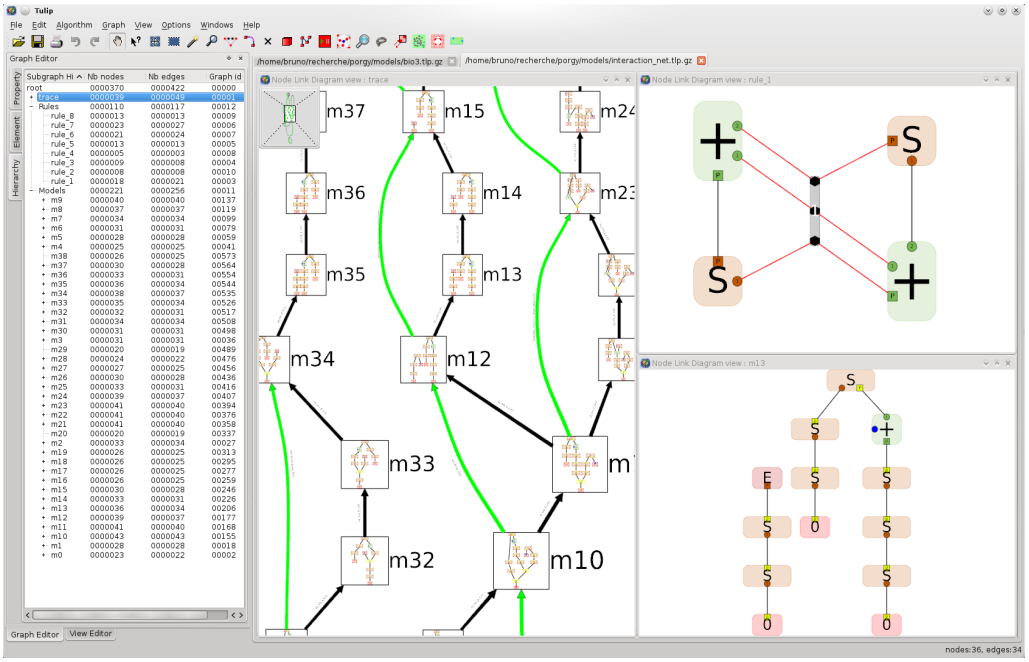

Figure 10. Partial view of the PORGY software application supporting visual exploration of graph rewriting systems. The graph on the middle of the figure is the trace tree which contains all rules applications. The green edges symbolize the result of a strategy written by the user.

8.3. ERDF RAUDIN

Participant: Maylis Delest.

- Project: RAUDIN (Recherches Aquitaines sur les Usages pour le Développement des dispositifs Numériques)
- Call: ERDF European Regional Development Fund and Région Aquitaine (FR)
- start/end December 2008 – December 2011

The project aims at gaining a better understanding on how resources are published and disseminated on the web, seen as a central knowledge diffusion mechanism in modern organizations. The acquired knowledge will serve organizations and institutions measures their visibility on the web. Our team is planned in a second phase of the project after data will be collected. Obviously the supervision of the publication process will build into a networks of various entities. The analysis of these networks through adequate representations and interactions should help the project experts.

8.4. ANR Jeunes Chercheurs EVIDEN

Participants: David Auber, Romain Bourqui, Jonathan Dubois, Guy Melançon, Bruno Pinaud.

- Project: EVIDEN (Exploration and Visualization of Dynamic Relational Data)
- Call: ANR JCJC 2010
- start/end December 2010 – November 2014
- Budget: 243,017 euros

In the EVIDEN project, we are interested in an emerging area in Information Visualization which deals with the exploration and the visual analysis of dynamic data. Our objective is to devise methods and algorithms for the visualization and navigation of dynamic and relational data. Following the visualization "pipeline" (see 1), research in the EVIDEN project focuses on four main topics:

1. The definition of a data structure that is versatile, flexible and optimized enough to store large dynamic and relational data. This data structure must be able to guarantee efficient access and update times.
2. The design of methods for the decomposition and extraction of regions of interest in dynamic data. Decomposition methods must, on one hand, group similar elements together and, on the other limit changes during data evolution. Methods to extract regions of interest must allow the detection and the extraction of sub-networks with atypical behaviors.
3. The design of efficient methods for the visualization of dynamic data: we focus on two main topics, the visual representation of the dynamic data and the visual representation of data evolution.
4. The design of interaction methods for dynamic data: adapting and/or formally redefining interaction methods for static data as in previous work is required by data evolution and dynamism.

The types of problems that the EVIDEN project is addressing emerge naturally in the application domains of bioinformatics and biology. Ultimately, it aims at contributing to the Health sciences. Other application domains could also benefit from these results.

9. Dissemination

9.1. Program committees and related activities

- Our team has organized the International Conference Eurovis 2010 last june.

- Members of GRAVITÉ are in charge of courses on Information Visualization as part of Université Bordeaux I Master's degree in Computer Science (Master recherche).

GRAVITÉ's members have been and will be involved in program or reviewing committees of major international and national scientific events and journals, such as:

- SYNASC Yearly Conference, Timisoara, Romania ((**PC**)G. Melançon, since 2005)
- CGIV Computer Graphics, Imaging, Visualization ((**PC**) G. Melançon, since 2005)
- ISVC International Symposium on Visual Computing ((**PC**) G. Melançon, since 2010)
- ACM Symposium on Software Visualization SoftVis 2006 and SoftVis 2008 ((**PC**) David Auber)
- Eurographics: IEEE EuroVis Symposium on Visualization ((**PC**)D. Archambault 2008, (**PC, Poster-Chair 2010**) D. Auber 2008..2010, (**PC, PC-Chair 2010**)G. Melançon 2009-2010)
- IEEE Infovis: Information Visualization International Conference ((**PC**)Guy Melançon, since 2001; (**PC**)Dan Archambault 2009, (**EXT**) David Auber 2006..2010, (**EXT**) Romain Bourqui 2008..2010)
- Graph Drawing International Conference ((**PC**)David Auber 2010, (**EXT**) Bruno Pinaud 2010, (**EXT**) Romain Bourqui 2010)
- IEEE PacificVis Conference ((**PC**) David Auber 2010, (**EXT**) Romain Bourqui 2009..2010, (**EXT**) Guy Melançon 2010)
- IEEE TVCG: Transaction On Visualisation and Computer Graphics ((**EXT**) David Auber 2008..2010, (**EXT**) Romain Bourqui 2010), (**EXT**) Guy Melançon 2010)
- Graphic Forum Journal ((**EXT**) David Auber 2008..2010)
- Journal of VIsual Languages and Computing (**EXT**) Guy Melançon 2010)

We also take part in various project evaluation process:

- Reviewer for the ANR Massive Data program 2006 – 2009 / 2007 – 2010
- President of the ANR Massive Data program 2006 – 2009 (Maylis Delest)
- External Reviewer for the ANR bi-lateral project Japan-France (David Auber 2009)
- Jury member for the ANR Massive Data program 2008 – 2011 (Guy Melançon)
- Jury member of the ASTI PhD Award, see http://www.asti.asso.fr/prix_these_2009 (G. Melançon)
- Jury member of the "La Recherche" magazine Award (G. Melançon)
- Jury member for the ANR - Contint 2010 (Guy Melançon)
- Jury member for the ANR - Blanc 2010 (Guy Melançon)
- Jury member for the ANR - COSCG 2010 (Guy Melançon)
- Jury member for the CNRS - PEPS 2010 (Guy Melançon)
- Jury member for the NWO (Netherland) - NSF USA 2010 (Guy Melançon)
- expert for Digiteo - Ile de France (Guy Melançon)
- expert for AAP interne INRA Loire (Guy Melançon)
- expert for AAP Université Paul Sabatier Toulouse (Guy Melançon)
- expert for Morgan Kaufmann (editor) (Guy Melançon)
- expert for BQR 2010 - Université Paris 13 - BRED (David Auber)

10. Bibliography

Major publications by the team in recent years

[1] G. ANDRIENKO, J.-D. FEKETE, C. GÖRG, D. KEIM, J. KOHLHAMMER, G. MELANÇON. *Visual Analytics: Definition, Process, and Challenges*, in "Information Visualization Human-Centered Issues and Perspectives", A. KERREN, J. T. STASKO, J.-D. FEKETE, C. NORTH (editors), LNCS State-of-the-Art Survey, Springer, 2008, vol. 4950, p. 154-175, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272779/en/.

[2] D. ARCHAMBAULT, T. MUNZNER, D. AUBER. *GrouseFlocks: Steerable Exploration of Graph Hierarchy Space*, in "IEEE Transactions on Visualization and Computer Graphics", 2008, vol. 14, p. 900–913, http://hal.inria.fr/inria-00338627/en/.

[3] D. ARCHAMBAULT, H. PURCHASE, B. PINAUD. *The readability of Path-Preserving Clusterings of Graphs*, in "Eurovis 2010, 12th annual Eurographics/IEEE Symposium on Visualization", France Bordeaux, G. MELANÇON, T. MUNZNER, D. WEISKOPF (editors), Eurographics/ IEEE-VGTC Symposium on Visualization 2010, WILEY, June 2010, vol. 29(3), p. 1173–1182 [*DOI :* 10.1111/J.1467-8659.2009.01683.X], http://hal.inria.fr/inria-00471432/en.

[4] D. AUBER, M. DELEST, J.-P. DOMENGER, S. DULUCQ. *Efficient drawing and comparison of RNA secondary structure, Journal of Graph Algorithms and Applications*, in "Journal of Graph Algorithms and Applications", 2006, vol. 10, n^{o} 2, p. 329-351.

[5] R. BOURQUI, L. COTTRET, V. LACROIX, D. AUBER, P. MARY, M.-F. SAGOT, F. JOURDAN. *Metabolic network visualization eliminating node redundance and preserving metabolic pathways*, in "BMC Systems Biology", 2007, vol. 1, n^{o} 29.

[6] M. DELEST, J.-M. FÉDOU, G. MELANÇON. *A Quality Measure for Multi-Level Community Structure*, in "Algorithms for Scientific Computing (SYNASC)", Timisoara (Romania), IEEE Computer Society, 2006, p. 63-68.

[7] PIERRE-YVES. KOENIG, F. ZAIDI, D. ARCHAMBAULT. *Interactive Searching and Visualization of Patterns in Attributed Graphs.*, in "Proceedings of Graphics Interface", Canada, 2010, p. 113–120, http://hal.inria.fr/hal-00499430/en.

[8] A. LAMBERT, R. BOURQUI, D. AUBER. *Winding Roads: Routing edges into bundles*, in "Computer Graphics Forum", June 2010, vol. 29, n^{o} 3, p. 853-862, http://hal.inria.fr/hal-00495279/en.

[9] A. TELEA, D. AUBER. *Code Flows: Visualizing Structural Evolution of Source Code*, in "Computer Graphics Forum", 2008, p. 831-938, http://hal.inria.fr/inria-00338601/en/.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] F. ZAIDI. *Analyse, Structure et Organisation des Réseaux Complexes*, Université Sciences et Technologies - Bordeaux I, November 2010, http://hal.inria.fr/tel-00542703/en.

Articles in International Peer-Reviewed Journal

[11] D. ARCHAMBAULT, T. MUNZNER, D. AUBER. *Tugging Graphs Faster: Efficiently Modifying Path-Preserving Hierarchies for Browsing Paths*, in "IEEE Transactions on Visualization and Computer Graphics", 2010, http://hal.inria.fr/inria-00413861/en.

[12] D. ARCHAMBAULT, H. PURCHASE, B. PINAUD. *Animation, Small Multiples, and the Effect of Mental Map Preservation in Dynamic Graphs*, in "IEEE Transactions on Visualization and Computer Graphics", 2010, vol. 17, To Appear, http://doi.ieeecomputersociety.org/10.1109/TVCG.2010.78, http://hal.inria.fr/inria-00472423/en.

[13] D. DERRICK, D. ARCHAMBAULT. *Treeform: Explaining and exploring grammar through syntax trees.*, in "Literary and Linguistic Computing", 2010, http://hal.inria.fr/inria-00413849/en.

[14] C. DUCRUET, C. ROZENBLAT, F. ZAIDI. *Ports in multi-level maritime networks: Evidence from the Atlantic (1996-2006)*, in "Journal of Transport Geography", 2010, vol. 18, n^{o} 4, p. 508-518 [*DOI :* 10.1016/J.JTRANGEO.2010.03.005], http://hal.inria.fr/halshs-00466163/en.

[15] F. GILBERT, P. SIMONETTO, F. ZAIDI, F. JOURDAN, R. BOURQUI. *Communities and hierarchical structures in dynamic social networks*, in "Social Network Analysis and Mining", September 2010, http://hal.inria.fr/hal-00520936/en.

[16] A. LAMBERT, R. BOURQUI, D. AUBER. *Winding Roads: Routing edges into bundles*, in "Computer Graphics Forum", June 2010, vol. 29, n^{o} 3, p. 853-862, http://hal.inria.fr/hal-00495279/en.

International Peer-Reviewed Conference/Proceedings

[17] D. ARCHAMBAULT, H. PURCHASE, B. PINAUD. *Difference Map Readability for Dynamic Graphs*, in "18th International Symposium on Graph Drawing", Allemagne Constance, LNCS, Springer, 2010, vol. 6502, To appear, http://hal.inria.fr/inria-00514150/en.

[18] D. ARCHAMBAULT, H. PURCHASE, B. PINAUD. *The readability of Path-Preserving Clusterings of Graphs*, in "Eurovis 2010, 12th annual Eurographics/IEEE Symposium on Visualization", France Bordeaux, G. MELANÇON, T. MUNZNER, D. WEISKOPF (editors), Eurographics/ IEEE-VGTC Symposium on Visualization 2010, WILEY, June 2010, vol. 29(3), p. 1173–1182 [*DOI :* 10.1111/J.1467-8659.2009.01683.X], http://hal.inria.fr/inria-00471432/en.

[19] U. BRANDES, S. CORNELSEN, B. PAMPEL, A. SALLABERRY. *Blocks of Hypergraphs applied to Hypergraphs and Outerplanarity*, in "21st International Workshop on Combinatorial Algorithms (IWOCA 2010)", Royaume-Uni, July 2010, to appear, http://hal.inria.fr/hal-00539143/en.

[20] U. BRANDES, S. CORNELSEN, B. PAMPEL, A. SALLABERRY. *Path-Based Supports for Hypergraphs*, in "21st International Workshop on Combinatorial Algorithms (IWOCA 2010)", Royaume-Uni, July 2010, to appear, http://hal.inria.fr/hal-00538984/en.

[21] F. GILBERT, D. AUBER. *From Databases to Graph Visualization*, in "2010 14th International Conference Information Visualisation", Royaume-Uni London, July 2010, 128, http://hal.inria.fr/hal-00520906/en.

[22] PIERRE-YVES. KOENIG, F. ZAIDI, D. ARCHAMBAULT. *Interactive Searching and Visualization of Patterns in Attributed Graphs.*, in "Proceedings of Graphics Interface", Canada, 2010, p. 113–120, http://hal.inria.fr/hal-00499430/en.

[23] A. LAMBERT, D. AUBER, G. MELANÇON. *Living flows: enhanced exploration of edge-bundled graphs based on GPU-intensive edge rendering*, in "14th International Conference on Information Visualization (IV'10)", Royaume-Uni, July 2010, p. 523-530, http://hal.inria.fr/hal-00495293/en.

[24] A. LAMBERT, R. BOURQUI, D. AUBER. *3D Edge Bundling for Geographical Data Visualization*, in "14th International Conference on Information Visualization (IV'10)", Royaume-Uni, July 2010, p. 329-335, http://hal.inria.fr/hal-00495307/en.

[25] B. PINAUD, P. KUNTZ. *GVSR: an On-Line Guide for Choosing a Graph Visualization Software*, in "18th International Symposium on Graph Drawing", Allemagne Constance, U. BRANDES, S. CORNELSEN (editors), LNCS, Springer, September 2010, vol. 6502, To appear, http://hal.inria.fr/inria-00516580/en.

[26] A. SALLABERRY, N. PECHEUR, S. BRINGAY, M. ROCHE, M. TEISSEIRE. *Discovering Novelty in Gene Data : From Sequential Patterns to Visualization*, in "6th International Symposium on Visual Computing (ISVC10)", USA, November 2010, p. 534-543, http://hal.inria.fr/hal-00539154/en.

[27] A. SALLABERRY, F. ZAIDI, C. PICH, G. MELANÇON. *Interactive Visualization and Navigation of Web Search Results Revealing Community Structures and Bridges*, in "Proceedings of Graphics Interface", Canada Ottawa, 2010, p. 105–112, http://hal.inria.fr/inria-00538564/en.

[28] F. ZAIDI, G. MELANÇON. *Evaluating the Quality of Clustering Algorithms using Cluster Path Lengths*, in "10th Industrial Conference, ICDM", Allemagne Berlin, 2010, http://hal.inria.fr/inria-00542690/en.

[29] F. ZAIDI, G. MELANÇON. *Identifying the Presence of Communities in Complex Networks Through Topological Decomposition and Component Densities*, in "EGC 2010, Extraction et Gestion de Connaissance", Tunisie Yasmine Hamamat, 163-174, 2010, vol. E-19, RNTI, p. 163-174, http://hal.inria.fr/inria-00538566/en.

[30] F. ZAIDI, G. MELANÇON. *Organization of Information for the Web using Hierarchical Fuzzy Clustering Algorithm based on Co-Occurrence Networks*, in "WI-IAT '10", Canada Toronto, 2010, http://hal.inria.fr/inria-00542691/en.

National Peer-Reviewed Conference/Proceedings

[31] D. AUBER, P. MARY, M. MATHIAUT, J. DUBOIS, A. LAMBERT, D. ARCHAMBAULT, R. BOURQUI, B. PINAUD, M. DELEST, G. MELANÇON. *Tulip: a Scalable Graph Visualization Framework*, in "Extraction et Gestion des Connaissances (EGC) 2010", Tunisie Hammamet, S. B. YAHIA, J.-M. PETIT (editors), RNTI, January 2010, vol. RNTI E-19, p. 623-624, http://hal.inria.fr/inria-00482563/en.

[32] A. SALLABERRY, N. PECHEUR, S. BRINGAY, M. ROCHE, M. TEISSEIRE. *SequencesViewer : comment rendre accessible des motifs séquentiels de gènes trop nombreux ?*, in "10e Conférence Internationale Francophone sur l'Extraction des Connaissances (EGC 2010)", Tunisie Hammamet, January 2010, p. 387-392, http://hal.inria.fr/hal-00538972/en.

Workshops without Proceedings

[33] C. BLANC, M. DELEST, J.-M. FÉDOU, G. MELANÇON, F. QUEYROI. *Évaluer la qualité d'une fragmentation de graphe multi-niveaux*, in "Journées MARAMI 2010", France Toulouse, October 2010, 12 pages, http://hal.inria.fr/hal-00542484/en.

[34] B. PINAUD, G. MELANÇON. *PORGY : réécriture et visualisation de graphes dynamiques*, in "Extraction et Gestion des Connaissances (EGC 2010), 8e Atelier Visualisation et Extraction de Connaissances", Tunisie Hammamet, January 2010, http://hal.inria.fr/inria-00449745/en.

Scientific Books (or Scientific Book chapters)

[35] R. BOURQUI, P. SIMONETTO, F. JOURDAN. *A stable decomposition algorithm for dynamical social network analysis*, in "Advances in Knowledge Discovery and Management", F. GUILLET, G. RITSCHARD, D. A. ZIGHED, H. BRIAND (editors), Studies in Computational Intelligence, Springer, 2010, p. 167-178 [*DOI :* 10.1007/978-3-642-00580-0_10], http://hal.inria.fr/hal-00543175/en.

References in notes

[36] C. JOHNSON, R. MOORHEAD, T. MUNZNER, H. PFISTER, P. RHEINGANS, T. S. YOO (editors). *NIH-NSF Visualization Research Challenges Report*, IEEE Computer Society, 2006.

[37] M. KAUFMANN, D. WAGNER (editors). *Drawing Graphs: Methods and Models*, LNCS, Springer, 2001, vol. 2025.

[38] J. J. THOMAS, K. A. COOK (editors). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE Computer Society, 2006.

[39] M. AMIEL, G. MELANÇON, C. ROZENBLAT. *Réseaux multi-niveaux : l'exemple des échanges aériens mondiaux*, in "M@ppemonde", 2005, vol. 79, n^{o} 3-2005, http://mappemonde.mgm.fr/num7/index.html.

[40] D. AUBER. *Tulip*, in "9th International Symposium on Graph Drawing, GD 2001", Vienna, Austria, P. MUTZEL, M. JÜNGER, S. LEIPERT (editors), Lecture Notes in Computer Science, Springer-Verlag, 2001, vol. 2265, p. 335-337.

[41] D. AUBER. *Using Strahler Numbers for Real-Time Visual Exploration of Huge Graphs*, in "International Conference on Computer Vision and Graphics", Journal of WSCG, 2002, vol. 10/1–3, p. 56–69.

[42] D. AUBER. *Tulip - A huge graph visualization framework*, in "Graph Drawing Software", P. MUTZEL, M. JÜNGER (editors), Mathematics and Visualization Series, Springer Verlag, 2003, p. 105-126.

[43] D. AUBER, Y. CHIRICOTA, F. JOURDAN, G. MELANÇON. *Multiscale navigation of Small World Networks*, in "IEEE Symposium on Information Visualisation", Seattle, GA, USA, IEEE Computer Science Press, 2003, p. 75-81.

[44] D. AUBER, M. DELEST. *A Clustering Algorithm for Huge Trees*, in "Advances in Applied Mathematics", 2003, vol. 31, n^{o} 1, p. 46-60.

[45] D. AUBER, M. DELEST, J.-P. DOMENGER, P. FERRARO, R. STRANDH. *EVAT: an Environment for the Visualization and Analysis of Trees (2nd place - InfoVis Contest)*, in "IEEE Symposium on Information Visaulization", IEEE Computer Society, 2003, p. 124-126.

[46] D. AUBER, P. MARY. *Mise en place d'un mécanisme de plugins en C++*, in "Programmation sous linux", 2006, vol. 5, n^{o} 5, p. 74–79.

[47] G. BAILLY, D. AUBER, L. NIGAY. *2m: un espace de conception pour i'interaction bi-manuelle*, in "UbiMob '05: 2nd French-speaking conference on Mobility and uibquity computing", ACM Press, 2005, p. 177-184.

[48] G. BAILLY, D. AUBER, L. NIGAY. *From visualization to manipulation of rna secondary and tertiary structures*, in "10th International Conference on Information Visualisation (IV'06)", IEEE Computer Society, 2006, p. 107-116.

[49] G. BAILLY, D. AUBER, L. NIGAY. *Navrna: Visualization - exploration - edition of RNA*, in "International Conference on Advanced Visual Interfaces (AVI '06)", ACM, 2006, p. 504-507.

[50] M. BALDONADO, Q. WANG, A. WOODRUFF, A. KUCHINSKY. *Guidelines for Using Multiple Views in Information Visualization*, in "Advanced Visual Interfaces", 2000, p. 110-119.

[51] V. BATAGELJ, A. MRVAR. *Pajek - Analysis and Visualization of Large Networks*, in "Graph Drawing : 9th International Symposium, GD 2001", Vienna, Austria, Lecture Notes in Computer Science, 2002, vol. 2265, 477.

[52] C. BOHAN, B. GAUTIER, C. ROZENBLAT, P.-Y. KOENIG. *Networks Of Urban Centers And Of Multinational Corporations*, in "15th European Colloquium on Theoretical and Quantitative Geography", Zurich, CH, 2007, p. 33–40.

[53] R. BOULET, B. JOUVE, F. ROSSI, N. VILLA. *Batch kernel SOM and related Laplacian methods for social network analysis*, in "NeuroComputing Special Issue on Progress in Modeling, Theory, and Application of Computational Intelligenc - 15th European Symposium on Artificial Neural Networks 2007", 2008, vol. 71, n^{o} 7-9, 12571273.

[54] S. BOZZANI-FRANC, A. CONESA, A. L'HOSTIS, D. AUBER, C. ROZENBLAT, P. MARY, G. MELANÇON, P.-Y. KOENIG. *Relating the organization of the cities network and the organization of the air transport network: an approach through graph visualization*, in "15th European Colloquium on Theoretical and Quantitative Geography", Zurich, CH, 2007, p. 47–51.

[55] E. H. CHI. *A Taxonomy of Visualization Techniques Using the Data State Reference Model*, in "IEEE Symposium on Information Vizualization", IEEE Computer Society, 2000, 69.

[56] E. H. CHI, J. RIEDL. *An Operator Interaction Framework for Visualization Systems*, in "IEEE Symposium on Information Visualization (InfoVis '98)", IEEE Computer Society, 1998, p. 63-70.

[57] Y. CHIRICOTA, G. MELANÇON, T. T. PHAN QUANG, P. TISSANDIER. *Visual Exploration of (French) Commuter Networks*, in "Geovisualization of Dynamics, Movement, and Change AGILE'08 Satellite Workshop", Espagne, 2008, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272786/en/.

[58] M. CHUAH, S. ROTH. *On the Semantics of Interactive Visualizations*, in "IEEE International SYmposium on Information Visualization '96", IEEE Computer Society, 1996, p. 29-36.

[59] M. DELEST, T. MUNZNER, D. AUBER, J.-P. DOMENGER. *Exploring InfoVis Publication History with Tulip (2nd place - InfoVis Contest)*, in "IEEE Symposium on Information Visaulization", IEEE Computer Society, 2004, 110.

[60] J. ELLSON, E. R. GANSNER, L. KOUTSOFIOS, S. C. NORTH, G. WOODHULL. *Graphviz - Open Source Graph Drawing Tools*, in "Graph Drawing : 9th International Symposium, GD 2001", Vienna, Austria, Lecture Notes in Computer Science, 2002, vol. 2265, 483.

[61] J. ELLSON, E. R. GANSNER, L. KOUTSOFIOS, S. C. NORTH, G. WOODHULL. *Graphviz and Dynagraph - Static and Dynamic Graph Drawing Tools*, in "Graph Drawing Software", M. JÜNGER, P. MUTZEL (editors), Springer-Verlag, 2003, p. 127-148.

[62] J.-D. FEKETE. *The InfoVis Toolkit*, in "10th IEEE Symposium on Information Visualization", Austin, TX, IEEE Press, 2004, p. 167–174.

[63] M. C. FERREIRA DE OLIVEIRA, H. LEVKOWITZ. *From visual data exploration to visual data mining: a survey*, in "IEEE Transactions on Visualization and Computer Graphics", 2003, vol. 9, n^{o} 3, p. 378-394, 1077-2626.

[64] G. GAINANT, D. AUBER. *Arna: Interactive comparison and alignment of rna secondary structure*, in "IEEE Symposition on Information Visualisation", IEEE Computer Society, 2004, 8.

[65] E. R. GANSNER, S. C. NORTH. *An open graph visualization system and its applications to software engineering*, in "Software – Practice and Experience", 2000, vol. 30, n^{o} 11, p. 1203-1233.

[66] D. J. HAND. *Strength in Diversity: The Advance of Data Analysis*, in "Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD 2004", Pisa, Italy, J.-F. BOULICAUT, F. ESPOSITO, F. GIANOTTI, D. PEDRESCHI (editors), Lecture Notes in Artificial Intelligence, Springer, 2004, vol. 3202, p. 18-26.

[67] D. J. HAND. *Statistics and data mining: intersecting disciplines*, in "ACM SIGKDD Explorations Newsletter", 1999, vol. 1, n^{o} 1, p. 16-19.

[68] J. HEER, S. K. CARD, J. A. LANDAY. *prefuse: a toolkit for interactive information visualization*, in "SIGCHI conference on Human factors in computing systems", Portland, Oregon, USA, ACM Press, 2005, p. 421-430.

[69] I. HERMAN, M. S. MARSHALL, G. MELANÇON. *Graph Visualisation and Navigation in Information Visualisation: A Survey*, in "IEEE Transactions on Visualization and Computer Graphics", 2000, vol. 6, n^{o} 1, p. 24-43.

[70] I. HERMAN, G. MELANÇON, M. DELEST. *Tree Visualisation and Navigation Clues for Information Visualisation*, in "Computer Graphics Forum", 1998, vol. 17, n^{o} 2, p. 153–165.

[71] INRIA. *Plan stratégique 2003-2007*, 2003.

[72] F. IRAGNE, M. NIKOLSKI, B. MATHIEU, D. AUBER, D. SHERMAN. *Proviz: protein interaction visualization and exploration*, in "Bioinformatics", 2005, vol. 21, n^{o} 2, p. 272-274.

[73] G. JAESCHKE, M. LEISSLER, M. HEMMJE. *Modeling Interactive, 3-Dimensional Information Visualizations Supporting Information Seeking Behaviors*, Lecture Notes in Computer Science : Knowledge and Information Visualization, Springer, 2005.

[74] D. KEIM, H.-P. KRIEGEL. *Visualization Techniques for Mining Large Databases: A Comparison*, in "IEEE Transactions on Knowledge and Data Engineering", 1996, vol. 8, n^{o} 6, p. 923-936.

[75] P.-Y. KOENIG, G. MELANÇON. *Combining DagMaps and Sugiyama Layout for the Navigation of Hierarchical Data*, in "IV '07: Proceedings of the 11th International Conference Information Visualization", Zurich, CH, IEEE Computer Society, 2007, p. 447-452.

[76] P.-Y. KOENIG, G. MELANÇON. *Dagmap : exploration interactive de relations d'héritage*, in "Revue d'Intelligence Artificielle", 2008, vol. 22, p. 355-370, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272829/en/.

[77] G. LOHSE, K. BIOLSI, N. WALKER, H. RUETER. *A classification of visual representations*, in "Communications of the ACM", 1994, vol. 37, n^{o} 12, p. 36-49.

[78] B. S. MANJUNATH, J.-R. OHM, V. V. VASUDEVAN, A. YAMADA. *Color and Texture Descriptors*, in "IEEE Transactions on Circuits and Systems for Video Technology", 2001, vol. 11, n^{o} 6, p. 703-715.

[79] G. MELANÇON, Y. CHIRICOTA. *Visually mining relational data*, in "International Review on Computers and Software", 2007, p. 1-14.

[80] P. MUTZEL, M. JÜNGER. *Graph Drawing Software*, Springer, 2003.

[81] U. A. PEREGO. *The Power of DNA: Discovering Lost and Hidden Relationships. How DNA Analysis techniques are assisting in the great search for our ancestors.*, in "World Library and Information Congress: 71th IFLA General Conference and Council", Oslo, Norway, 2005, p. 1-19.

[82] L. PETER, H. R. VARIAN. *How Much Information*, 2003, n^{o} March 2006.

[83] H. E. ROMAN, M. ALBERGANTE, M. COLOMBO, F. CROCCOLO, F. MARINI, C. RICCARDI. *Modeling cross correlations within a many-assets market*, in "Physical Review E", 2006, vol. 73, n^{o} 3, 036129.

[84] C. ROZENBLAT, R. BOURQUI, D. AUBER, G. MELANÇON. *Community identification in social networks : a comparative survey*, in "SUNBELT XXVIII International Sunbelt Social Network Conference", INSNA, 2008.

[85] C. ROZENBLAT, P.-Y. KOENIG, G. MELANÇON, J.-C. LOUBIER. *Territorial and Topological Levels in Worldwide Air Transport Networks (2000-2004*, in "15th European Colloquium on Theoretical and Quantitative Geography", Zurich, CH, 2007, p. 365–371.

[86] B. SHNEIDERMAN. *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization*, in "IEEE Conference on Visual Languages (VL'96)", Boulder, CO, IEEE CS Press, 1996, p. 336-343.

[87] T. SIKORA. *The MPEG-7 Visual Standard for Content Description – An Overview*, in "IEEE Transactions on Circuits and Systems for Video Technology", 2001, vol. 11, n^{o} 6, p. 696-702.

[88] M. TORY, T. MÖLLER. *Rethinking Visualization: A High-Level Taxonomy*, in "IEEE Symposium on Information Visualization 2004", 2004, p. 151-158.

[89] M. O. WARD. *A taxonomy of glyph placement strategies for multidimensional data visualization*, in "Information Visualization (Palgrave)", 2002, vol. 1, n^o 3/4, p. 194-210.

[90] C. WARE. *Information Visualization: Perception for Design*, Morgan Kaufmann Publishers, Orlando, FL, 2000.

[91] D. J. WATTS, S. H. STROGATZ. *Collective dynamics of "small-world" networks*, in "Nature", 1998, vol. 393, p. 440–442.

[92] D. J. WATTS. *Six Degrees: The Science of a Connected Age*, 1st, W. W. Norton & Company, 2003.

[93] D. J. WATTS. *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, 1999.

[94] W. DE NOOY, A. MRVAR, V. BATAGELJ. *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, 2004.

[95] S. DOS SANTOS, K. BRODLIE. *Gaining understanding of multivariate and multidimensional data through visualization*, in "Computers & Graphics", 2004, vol. 28, n^o 3, p. 311-325.

[96] J. VAN WIJK. *The Value of Visualization*, in "IEEE Visualization", C. SILVA, E. GROELLER, H. RUSHMEIER (editors), IEEE Computer Society, 2005, p. 79-86.