



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team GRAVITÉ

*Graph Visualization and Interactive
Exploration*

Bordeaux - Sud-Ouest

THEME COG

Activity
R *eport*

2008

Table of contents

1. Team	1
2. Overall Objectives	1
2.1. Introduction	1
2.2. Highlights	2
3. Scientific Foundations	3
3.1. Introduction	3
3.2. Visual Graph Mining	4
3.3. Visual Analytics for Graphs	5
4. Application Domains	6
4.1. Introduction	6
4.2. Social sciences	6
4.3. Biology and bio-informatics	7
4.4. Strategic watch	8
5. Software	8
6. New Results	10
6.1. Graph Drawing	10
6.1.1. Drawability of Euler diagrams	10
6.1.2. Adaptive DAG drawing	12
6.2. Graph statistics and graph clustering	12
6.2.1. Graph comparison	12
6.2.2. Metric survey	13
6.3. Interactive graph mining	13
6.3.1. GrouseFlocks: Steerable Exploration of Graph Hierarchy Space	13
6.3.2. Code flows	16
6.3.3. Hierarchies	17
6.3.4. Applications	17
6.3.4.1. Community identification	17
6.3.4.2. Subsidiaries network	17
6.3.4.3. Commuter network	17
7. Contracts and Grants with Industry	20
8. Other Grants and Activities	21
8.1. ANR Systryp	21
8.2. ANR SPANGEO	22
8.3. ERDF RAUDIN	23
9. Dissemination	23
10. Bibliography	23

GRAVITE is an INRIA Team joint with University of Bordeaux (UBI) and CNRS (LaBRI, UMR 5800)

1. Team

Faculty Member

Guy Melançon [University Bordeaux 1, Team Leader, University Professor (Pr), HdR]
David Auber [University Bordeaux 1, Assistant Professor (MdC)]
Maylis Delest [University Bordeaux 1, University Professor (Pr), HdR]
Jean-Philippe Bordeaux Domenger [University Bordeaux 1, University Professor (Pr), HdR]
Bruno Pinaud [University Bordeaux 1, Assistant Professor]

Technical Staff

Jonathan Dubois [University Bordeaux 1, Research Engineer]
Antoine Lambert [University Bordeaux 1, Research Engineer]
Patrick Mary [University Bordeaux 1, Research Engineer]
Morgan Mathiaut [INRIA, Research Engineer]

PhD Student

Romain Bourqui [Université Bordeaux 1, Conseil Régional d'Aquitaine grant, since Apr. 2007]
Frédéric Gilbert [University Bordeaux 1, MENRT grant, since Oct. 2008]
Pierre-Yves Koenig [University Montpellier 2, Research contract, since Apr. 2007]
Trung Tien Phan Quang [University Bordeaux 1, MENRT grant, since Apr. 2007]
Arnaud Sallaberry [PIKKO CIFRE, since Mar. 2008]
Paolo Simonetto [University Bordeaux 1, MENRT grant, since Nov. 2007]
Faraz Zaidi [SFERE, since Sep. 2007]

Post-Doctoral Fellow

Daniel Archambault [INRIA, from September 1st, 2008 till September 30, 2009]

Visiting Scientist

Alexandru C. Telea [University of Groningen, Nederland, from January 1st till March 31, 2008]
Helen C. Purchase [University of Glasgow, United Kingdom, from April 7 till April 18, 2008]

Administrative Assistant

Alice Rivière [INRIA]

2. Overall Objectives

2.1. Introduction

A recent U.S. report to the funding agencies NIH and NSF [33] 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 [33]].

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 [75] [33] [34]. 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

01/2008 Guy Melançon was appointed as a member of the steering committee for the I3 journal (see the URL <http://www.revue-i3.org>). The Information - Interaction - Intelligence (I3) Journal was founded by members of the research group GDR I3, created by the CNRS in January 1998.

The journal serves as a forum favoring the dissemination of interdependent and interdisciplinary research involved in information retrieval, database, human-machine communication and artificial intelligence. The recent developments in these fields call for cross-discipline collaborations and perspectives. The I3 journal intends to promote inter-disciplinarity working at the development of common terminology, concepts, typologies, methods and ontologies and aiming at the identification of a unified perspective on research issues.

The journal is published both as a traditional paper publication (published by Cépaduès <http://www.cepadues.com/>) and as an electronic journal on the web.

04/2008 Our Graph Visualization framework was released under version Tulip 3.0 last April. One of the new features provided with this version is a *plugin manager* easing the download and update of the entire collection of plugins developed by the team through a server. Each time the Tulip GUI is launched, Tulip automatically connects to the plugin server hosted at LaBRI to check the availability of new plugins or application updates. Each connection is logged with the current date and client IP initiating the connection. Exploring the collection of user logs provided feedback on the actual use of the framework by our user community. Assuming that different IPs correspond to different users, excluding LaBRI IPs, the statistics showed that our 1600 users distribute over 69 different

countries. Based on user logs collected since the tracking system was launched, about 40 different users from 10 different countries launch the application everyday. Other statistics extracted from sourceforge.net (hosting the Tulip sources and binaries) and <http://www.tulip-software.org> (Tulip's website) also reveal the increasing evolution of our user community: each new version of Tulip is downloaded about 3300 times and this number has increased of 38% during the last twelve months period. Tulip's website is visited about 2500 times per month and this number has increased of 15% over the same period of time.

(07/2008) A subset of all our PhD students (and our intern Umang Sharan) took part in the VAST contest 2008 (mini challenge number 3, see the URL <http://www.cs.umd.edu/hcil/VASTchallenge08/>). The primary task of the mini challenge was to provide characteristics for the Catalano/Vidro social network and to provide evidence on the temporal changes occurring in this social structure. The data set gathered information on 9834 (mobile) phone calls between 400 cellphones over a 10 day period in June 2006 on the Isla Del Sueno. The records were expected to provide critical information about the Catalano social network structure. Given the quantity and nature of the data involved, our student team developed a visualization environment built on top of our framework Tulip [4] to help analyze the evolution of the call structure over time. The various relationships between members of the network were modeled as a *phone call* graph and a *location* graph. Temporal snapshots of the graphs were used as an analytic artefact to assist investigators analyzing the Catalano social network structure. The strength of the approach lied in its ease of use when exploring and navigating through the data set, easily switching between different views.

This work won our team a publication in the VisWeek Conference Compendium [22], [27].

(10/2008) GRAVITÉ will enjoy a three year grant from the ANR, working in collaboration with Thalès Communications and Xerox. The project called TANGUY – from Text to Arguments through Networks with Goals and User Initiative – will be concerned with the analysis of text fragments from which implicit pieces of information will be extracted, mixed with user data and interaction, focused at hypothesis building and validation ultimately helping law experts to build cases.

(11/2008) The SPANGEO ANR project in which we took part was elected as Highlight of the Masses de données 2005 call, thanks to its coordinator Professor Céline Rozenblat (now at University of Lausanne, Switzerland). The project gathered quantitative geographers and computer scientists in an effort to develop and exploit results from network science. The project had a decisive impact on the geographers ability to study data describing how network flows organize, confronting multi-level organizations of flows against territorial strategies.

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 [95], 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 [88]. 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 [70], [69] :

“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 Hands 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 [34] 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:

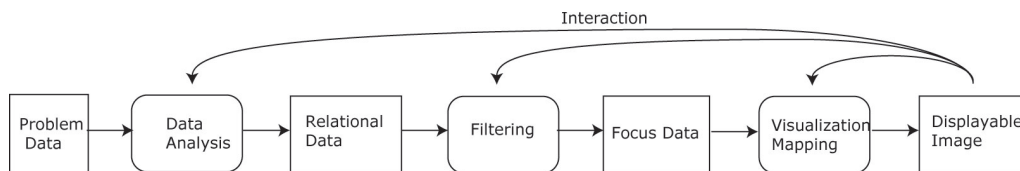


Figure 1. Visualization pipeline (adapted from [100]).

A decade ago, Ben Shneiderman¹ - 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” [90]. The pipeline is coherent with Shneidermans mantra which actually provides 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 analysis²:

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

Keims 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

¹Ben 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).

²See 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>

through Shneiderman and Keims 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 [102].

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 [82] [91] 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 Symposium³, the IEEE/Eurographics EuroVis Conference⁴ or the IEEE London Information Visualization Conference⁵ concern graph visualization. The Graph Drawing community, with its own annual international symposium also contributes to the development of the field⁶.

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

³See the URL <http://www.infovis.org>

⁴See the URL <http://www.eurovis.org>

⁵See the URL <http://www.graphicslink.co.uk>

⁶See the URL <http://www.graphdrawing.org>

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 [38]. Other examples from our group exploit the fact that combinatorial parameters in a tree can be approximated using a Gaussian distribution [74] [39], 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.” [33]

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 users 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.

4.2. 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.” [33]

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) project⁷. 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 [1] [51], [79], [49], [89] [26], [23], [16]. 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 [96], [98], [97] 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 [5] [11] [83] appears as a fruitful strategy to discover scales in datasets [1].

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.

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 project⁸. RNA comparisons required the development of new graph

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

⁸See the URL <http://www.lri.fr/~denise/AReNa/>

drawing algorithms and graph visualization software [65] [6]. 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 [8]. 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 [76].

The recently launched ANR project SysTryp already gave us the occasion to renew and 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. Strategic watch

All actors from the industry and from public domain 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. We actually work within the framework of a 2-year project funded by the ANR (RNTL program). This project is concerned with processing on-line web data (blogs, for instance). Our work builds on top of the AMI Intelligence software from AMI Software⁹ processing huge amounts of digital, numerical and textual data we still need to sort, classify, induce a topology on and cluster. The development of visualization software components is accomplished in collaboration with PIKKO Software¹⁰.

The goal is to propose incremental statistics and adjusting visualizations to support competitive and strategic watch. Ultimately, analysts want to be able to identify pieces of information acting first as outliers later confirming general trends. These pieces of information are what Ansoff called *weak signals* [36]. We are here out of the reach of classical statistics: analysts need to inject their knowledge and intuition in the system to help judge of anecdotic situation and put pieces of information under surveillance.

Our efforts focus on graph analysis and visualization, since most of the processed data can naturally be equipped with relations. The project requires the design of astute graph statistics and adaptative algorithms that can adjust with a highly changing environment. We revisited most of the existing work on text mining and document clustering, trying to exploit the scale free nature of the collected data.

We see in this project a real opportunity to address issues raised by the processing of on-line web data. In a sense, this project should reinforce our expertise on the processing and visualization of dynamic data. We already have visibility from the industry as shown by a recent invitation to participate to the Carto2.0 forum [24]. We also had the opportunity to present our work in the context of a NATO Series conference [29].

5. Software

5.1. Tulip: a scalable Graph Visualization Framework

Keywords: *Framework, Graph Library, Graph Visualization, OpenSource.*

⁹See the URL <http://www.amisw.com>

¹⁰See the URL <http://www.pikko-software.com>

Participants: David Auber [correspondant], Patrick Mary.

See also the web page <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 [4]. 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 [41]. We do however experiment with other information visualization platforms such as the InfoVis Toolkit developed at INRIA [60] and *prefuse* [72] 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 Trolltechs 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 [46] [99], [67] [58], [59] and others.

Tulips 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. Tulips 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 [41].

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 [40];
- SWViz implemented a specific hierarchical clustering algorithm for multilevel visualization and navigation of small world networks (social networks, software re-engineering graphs, etc.) [5];
- 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 [66];
- 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 [12].

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 [87], where graph clustering is astutely used to trace back genealogy from DNA data. Roman et al. [13] 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 [50] (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 [44] [43] [42]

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 [37] [4].

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 [99], for instance.

Part of Tulips 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 [81] [55] [78] [90] [54] [45] [53] [73] [94] [61] [93] [77].

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 service¹²) has already been created on sourceforge with that mid and long-term evolution of Tulip in mind.

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

6. New Results

6.1. Graph Drawing

Keywords: *Euler diagrams, Graph Drawing, overlapping subgraphs.*

Participants: David Auber, Paolo Simonetto, Bruno Pinaud.

6.1.1. Drawability of Euler diagrams

Visualization of taxonomies and other kinds of intersecting sets has recently become a focus of interest in the InfoVis community. The increasing number of application fields where cataloguing elements in strict hierarchies is no more enough. The need to take uncertainty and fuzzyness into account confirms the demand of visualisation methods that are able to handle these new structures. Euler diagrams, probably the most common and natural way of depicting intersecting sets, might provide the answer to these problems.

¹¹Eclipse 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.

¹²See the URL sourceforge.net/projects/tulip-plugin-ws/.

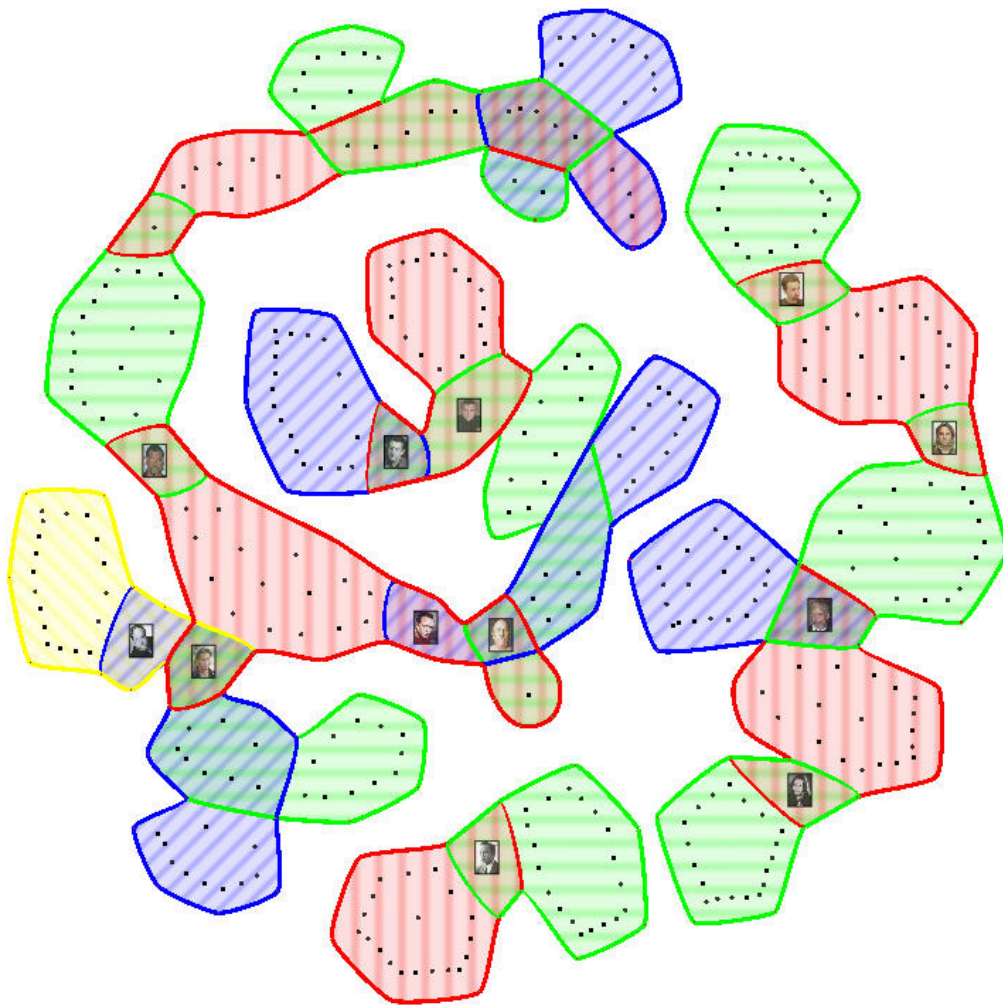


Figure 2. The diagram shows clearly how movies share their cast. Actors who take part in more than one movie are easily detectable in the overlapped regions of the graph. It is also evident how textures greatly help to recognize which films are sharing the same area, rather than a plain colour filling.

In Euler diagrams sets are identified by 2D regions enclosed by curves. The set intersections are then represented creating overlaps between the regions of the relative sets. This representation is so clear and intuitive that it is widely used in many different fields, also to introduce the mathematical concept of sets in elementary schools. Unfortunately, not all collection of sets admit a 2D Euler diagrams.

Euler representations are diagrams designed to extend the standard Euler ones, allowing to graphically depict even collection of intersecting sets otherwise undrawable [19]. They differ from standard Euler diagrams for the presence of visualisation patterns used when needed to overcome critical situations for standard Euler diagrams.

The automatic generation of these diagrams is possible through an algorithm described in [92] and implemented in Tulip. This algorithm differs from other approaches already developed not only for its capability of dealing with every input instance, but also for the usage of techniques not yet adopted for Euler diagrams. Among others, the depiction of the contours using Bézier curves and the application of transparent coloured textures for improving the readability of the overlapped regions.

Fig. 2 shows the Euler representation calculated for a collection of films. Each film is intended as a set of actors, and overlaps are present when the same actor is in the cast of two or more films.

The films chosen for the pictures are a highly interconnected subset of the IMDb top 40 films. The Internet Movie Database (IMDb) is an online database of information related to films, actors and more generally to the cinema world. It reports a chart containing the top 250 films, ranked according to an overall vote calculated on the viewer rates. For sake of readability, each film in the picture contains only the first 20 actors in credit order.

6.1.2. Adaptive DAG drawing

Industrial food processes often strongly rely on human expertise, expressed as know-how and control procedures based on subjective measurements (color, smell, texture), which are very difficult to capture and model. As a consequence, once the process is modeled as a graph its visualization can be used to support its study and supervision.

Algorithms previously designed in [80] were adapted to the case of cheese ripening, seen as a complex ecosystem and modeled as a DAG. The complexity of the process leads to uncertainty and require human experts to play a decisive role in the modeling process. As a consequence, the drawing procedure [25] [17] had to be adapted in order to deal with uncertainty and changing structures and attributes, ultimate helping to capture (learn) in a satisfying way expert knowledge.

6.2. Graph statistics and graph clustering

Keywords: *Network indices, centrality measures, correlation, edges metrics, node metrics, similarity measures.*

Participants: Guy Melançon, Arnaud Sallaberry, Maylis Delest.

Visual graph analytics definitely relies on the use of node and edge metrics to identify salient properties in graphs. Most often, these metrics are turned into useful visual cues, or are used to interactively filter out parts of a graph while querying it, for instance. Along the years, analysts coming from different application domains have designed metrics to serve specific needs. We report here our efforts to develop and study graph metrics as part of our research agenda.

6.2.1. Graph comparison

Comparing graphs remains a first order challenge because of its obvious and wide applicability. Because graphs may be large, or because one might not want to perform a rough comparison of graphs, visually comparing graphs may appear as a good trade-off solution. The exact comparison of graphs anyhow is intractable as it relies on the subgraph isomorphism problem. It turns out that many application domain only require the identification of quasi-similar graph patterns, allowing the design of efficient and lower complexity approaches.

The method we developed is based on a heuristic exploiting structural properties of graphs. Quasi-similar patterns are paired and colored accordingly. The method was first developed and used to compare trees to help locate differences in hierarchies [40]. The method was further adapted to compare and draw RNA secondary structures [6].

We recently extended and formalized the method to work with general graphs [20]. The heuristic is a two phase process. First, nodes are classified into different families (classes computed from node attributes). The second phase searches for quasi-similar patterns based on a labeling of neighbors. The covering tree introduced in [47] is used to optimize the construction of node classes. We moreover defined varying strategies to compare node neighborhoods.

6.2.2. Metric survey

Graph analytics, sometimes also called *network science*, recently developed as a cross-discipline field developing models shared by numerous application domains such as bio-informatics, social network analysis, web graphs, etc. [31], [56]. As a consequence, we end up finding various metrics in the literature aiming at similar goals; different names and analytics description often hide similarity between two metrics that originated from different fields.

Social network analysis grew out of Moreno's (see [62]) seminal work giving birth to sociometry. Later on, the development of graph theory by Harary (see [71] for instance) brought a decisive contribution to this field, establishing over the years a series of network metrics as a systematic method to tackle the study of social network structure. Network metrics provide a useful device to capture structure and identify or understand roles held by actors in a network (see [63], [32]). This is assessed by the continuously growing number of research papers relying on them when studying network of interacting entities. Sociometry, and more generally social network analysis, indeed extends to other research areas such as computer science and biology, for instance. Modules of source code interact by triggering each other's methods when a program executes. Proteins interacting in a cell form a network from which biologists hope to infer biological evidence.

As a consequence, researchers from various disciplines implicitly contribute ideas and methods to social network analysis. We started to survey a large set of measures defined on networks aiming at the identification of specific structure patterns [18]. Ad hoc measures that have emerged from independent sub-domains appear as being closely related to earlier ones and are compared on the basis of two different network models, scale-free versus (strictly) small-world. Our study aims at identifying "ancestor" metrics as significant archetypes for a number of recently introduced measures. We also seek for properties of each of the studied metric. This work is on-going. We also plan to look at how uncorrelated measures can be combined to help identify typical and generic roles of nodes in a scale-free or small world network.

6.3. Interactive graph mining

Participants: Daniel Archambault, David Auber, Romain Bourqui, Pierre-Yves Koenig, Guy Melançon.

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 whole Information Visualization community has undertaken a deep looking back at its achievements in order to decide of the next 10 years research priorities. We take part in this process in various form. Our presence at the Dagstuhl seminar [28] was one occasion we had to contribute with our own perspective on these issues. We also lately contributed to a collective monography in an effort to bring visualization closer to larger audience, as well as to specialists [30].

6.3.1. GrouseFlocks: Steerable Exploration of Graph Hierarchy Space

Grouse [3] enabled the interactive exploration of a large graph and associated hierarchy of topological features. In the project, metanodes, nodes containing subgraphs of the larger graph, could be opened or closed causing

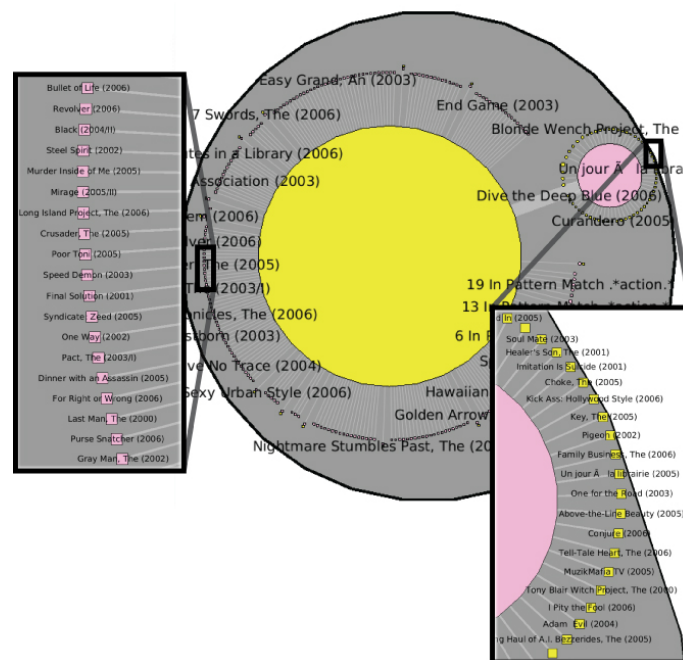


Figure 3. Decomposition into films that have been classified as part of the action genre, in pink, and those that are not, in yellow. Notice that there are two large components and many very small components in the graph. This indicates that if an actor acts in one action movie the actor tends to act with many other actors of the genre.

only part of the graph to be drawn on-demand by the user. This operation allowed for immediate exploration of very large graphs without the need for a precomputed drawing. These metanodes contained topological features, or patterns in connectivity between the nodes in the subgraph.

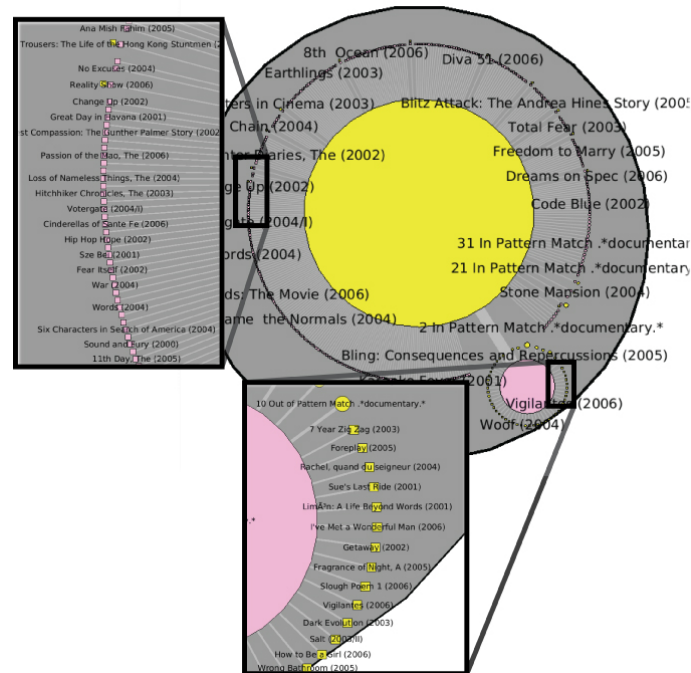


Figure 4. The same movie graph decomposed into movies that have been classified as part of the documentary genre, in pink, and those that are not in yellow. The trend continues where there are two large components and many small ones, indicating genre lock.

However, in most circumstances, attributes are associated with the nodes of the graph. These attributes can be used to construct a variety of hierarchies on top of the graph, providing different views of the underlying graph according to this information. For example, consider a computer networking dataset where nodes are servers and edges are connections between those servers. Attributes associated with the nodes of the graph group these servers into subnetworks. By grouping subnetworks together, we can reveal a higher level structure in the graph which is not supplied by the hierarchy of topological features. GrouseFlocks extends Grouse, as well as other steerable systems in previous work [57], [35], [85], to allow for the exploration of hierarchies based on attribute data.

With GrouseFlocks [14], users can create graph hierarchies interactively based on selections derived from the data associated with the nodes and edges of the graph, and Grouse is used to visualize the created hierarchies on the fly. The hierarchies generated are path-preserving, meaning that if a path exists in the abstract view of the graph, at least one path exists in the underlying graph. When no features are found and the size of the graph is above a user specified threshold, the system resorts to topological decomposition to reduce graph complexity. Visualization of graph hierarchies takes seconds while interactive creation of the hierarchies takes a couple of minutes. The approach was tested on graphs of tens of thousands of nodes and edges.

Figures 3, 4, and 5 are example hierarchies generated by GrouseFlocks. The underlying graph is a movie graph where nodes are movies and an edge exists between two movies if they share an actor. The generated

hierarchies indicate a tendency for genre lock, or when an actor acts in a particular genre of movie, the actor tends to act with other actors in that genre.

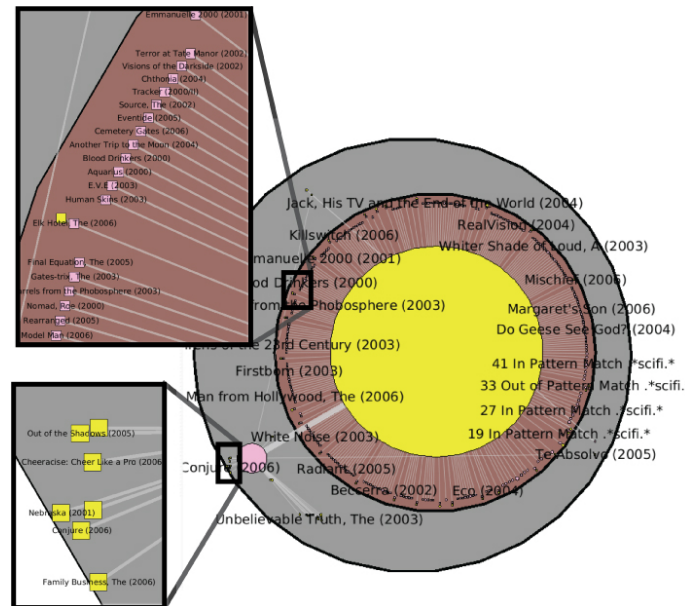


Figure 5. The same movie graph decomposed into movies that have been classified as part of the science fiction genre, in pink, and those that are not in yellow. The decomposition indicates genre lock.

6.3.2. Code flows

Analyzing the evolution of source code bases has become an important part of software maintenance. Code bases, stored in software configuration management systems such as CVS, Subversion or ClearCase, contain a wealth of evolution data, ranging from coarse-scale events, such as the addition or removal of a team member, to fine-scale events, such as the edits done to each line of code in every file. Several tools and techniques have been created to perform various types of analyses on code repositories. Such tools combine data mining components that extract the actual facts of interest from the repository with visualizations that let users see, navigate, and query the extracted facts to support specific tasks.

Recently, a technique was proposed to visualize code structure changes at a syntactic level, by rendering syntax-matched code blocks in consecutive versions as icicle plots, and connecting these with straight tubes [52]. We have extended this idea in several directions [15]. First, we render code correspondences using textured splines connected to mirrored icicle plots, thereby simplifying the visual inspection. Second, we present structure tracking, a technique to detect and display code fragments which stay (near) constant during evolution, and separate them from highly changing code. Finally, we describe how to detect and visualize complex events of interest, such as code splits and merges.

Many extensions of the code flows are possible. One of them is to show code clone evolution across several versions. Next, we plan to add information such as code quality metrics, programmer IDs, or bug data, mined from code repositories to the code flows, to assess e.g. how code marked as buggy drifted during a project, who was responsible for it, and thereby detect potential problems. This is helpful in both corrective and preventive maintenance.

6.3.3. Hierarchies

Hierarchies appear as a natural and important structures encoding information. Although they classically map to trees, they also encode inheritance relations where nodes have multiple direct ancestors (parents). These inheritance relations are described with acyclic directed graphs (DAGs).

We had the opportunity to study large hierarchies describing how world companies relates with their subsidiaries, and subsidiaries of these subsidiaries, and so on and so forth. Because classical node-link diagrams have limitations, we designed an extension of Treemaps to fit DAGs into these space-filling representations (or mosaic displays) [84]. The basic idea is to deploy a DAG into a tree by duplicating nodes with multiple ancestors. The dagmap as we call it, comes equipped with a set of possible interactions that help recover the DAG structure. Selecting a node, for instance, highlights its occurrences at various levels in the Treemap representation. The application of the dagmap within the SPANGEO project helped geographers identify and compare the strategies companies develop when developing part of their activity abroad [49].

The DAGMap was extended and equipped with specific interactors helping the user to select an adequate portion of the entire DAG under study [16], [21]. The interactor shown on Fig. 6 consists in selecting a *cut* of the DAG. The cut is determined by first computing an attribute for each node in the DAG, going bottom up from sink nodes to sources. The value associated to a node in a sense measures just how much its underlying DAG is homogeneous. This idea is borrowed from [101] where the authors deal with the particular case of a tree. In this case, the tree can be seen as the Hasse diagram of an ordered set and the cut coincides with a maximal antichain. With the DAG however, the cut no longer corresponds to an antichain since the DAG is no more a Hasse diagram of any ordered set.

The interactor let the user select a cut, either as a global view on the DAG, or as a contextual view where a focus node has been previously selected. This can be useful when focusing on a given element. Further elements are discarded and represented by higher level nodes while closer elements are kept as part of a detailed view.

6.3.4. Applications

6.3.4.1. Community identification

Graph hierarchies as a mean to visually explore and analyze network data have been applied to the study of the European and the world passenger air traffic. In [26], we reconsidered and surveyed methodologies originally studied in [51] [1], and tested the robustness of the method by varying both the graph statistics and the clustering schema used to infer communities. In each case, the statistics and quality criteria applied to cluster the graph lead to fine-grained differences between the resulting clustering, leading to a better understanding of both the statistics and the underlying data. This is on-going work taking place within the SPANGEO ANR project (in collaboration with CNRS UMR 5800 LaBRI).

6.3.4.2. Subsidiaries network

The DAGMap was designed in order to visualize hierarchies described as directed acyclic graphs (DAGs) [79]. However, DAGMap, just as the TreeMap they generalize, only behave well when the underlying hierarchy expands over a few levels. Deep hierarchies cause a recursive nesting of cells into extremely small portions of the screen, making them hard to read. As a consequence, the navigation of our 600K links hierarchy had to be equipped with a device selecting a relevant subpart of the hierarchy before feeding it into the DAGMap [21], [16]. The device was turned into an interactor mimicking the focus+context technique from Furnas [64] and extending the variation by van Ham and van Wijk [101].

6.3.4.3. Commuter network

Although inhabitants of greater cities generally work in the same town they live in, most non urban citizen commute to find work out of their hometown. As a consequence, governments must plan transportation and territorial accessibility accordingly, turning commuters into a first order subject of study for geographers. Commuters induce a network structure on the territory linking cities. Some cities appear as attractors because they offer good job opportunities, others appear as worker broadcasters because of their suburban activity (see Fig. 7).

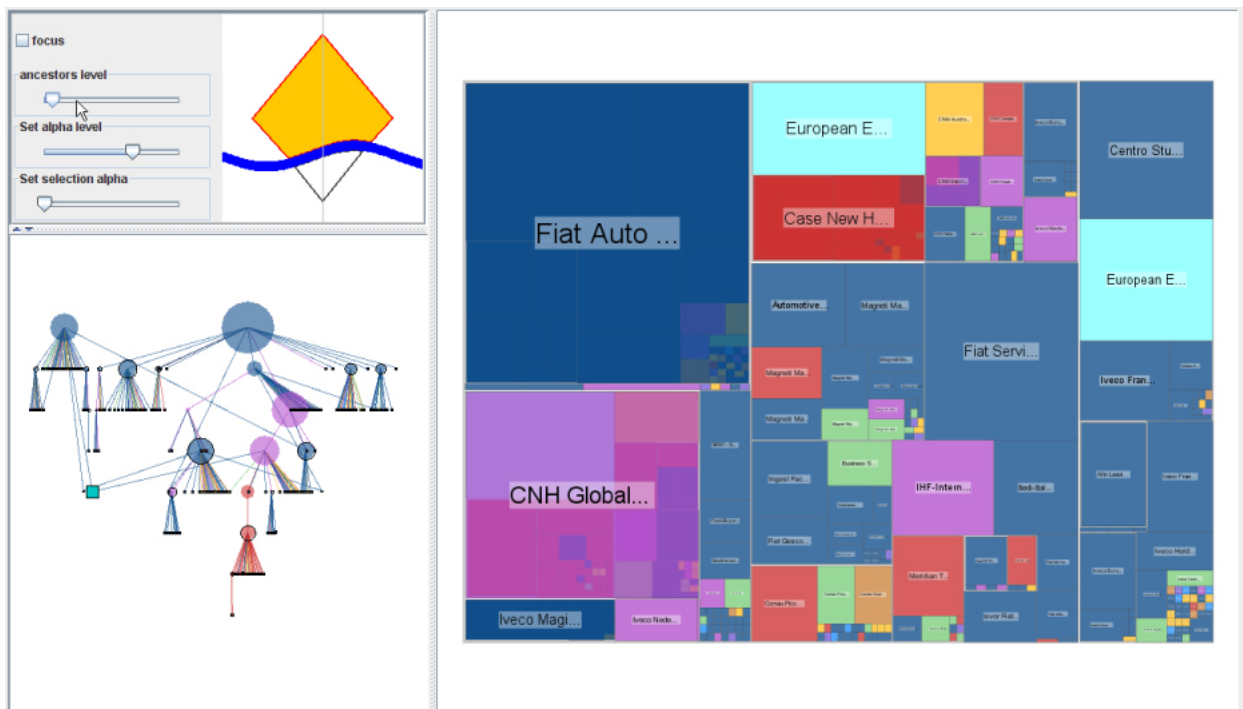


Figure 6. The DagMap comes equipped with adapted interaction to help gain details on lower levels and yet recover structure from upper levels. The diamond shaped interactor helps select a relevant portion of the whole datasets before feeding the corresponding DAGMap and node-link views.

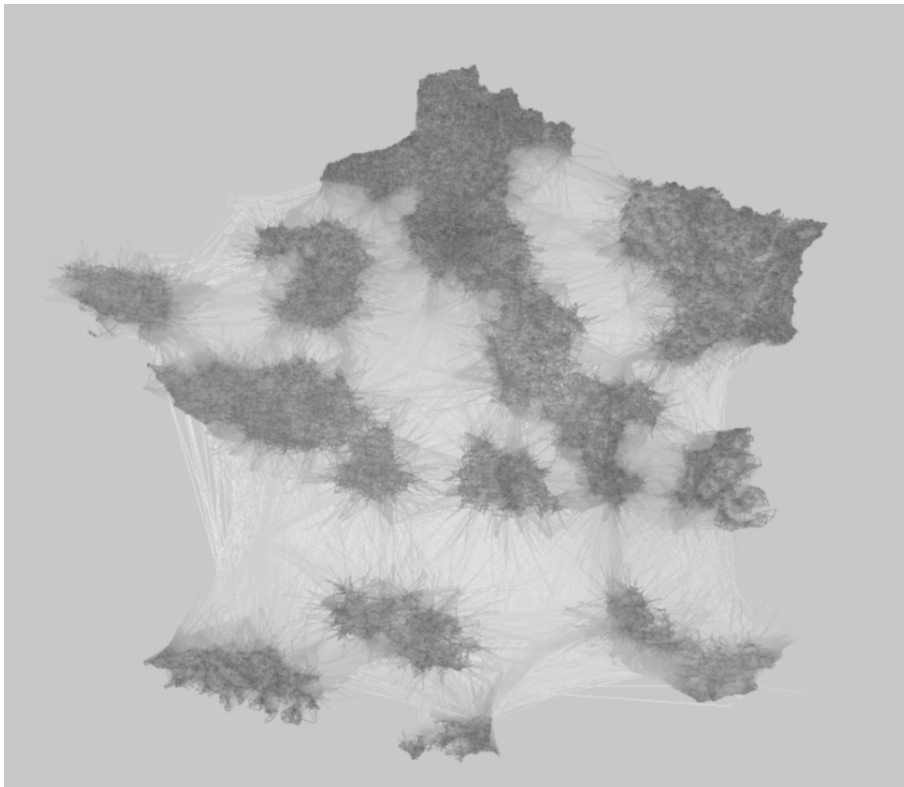


Figure 7. The commuter network as it forms over the French territory. Daily travels link cities. The underlying graph formed of cities (nodes) and daily travels (links) contains 33 thousand nodes and about 2 million links. Relevant interaction allows to filter links according to various metrics.

Most studies looking at commuter traffic have used simplified representations, such as egocentric networks concentrating on the exchange of workers going to the same city or region [48], [86]. Our approach [23] is more specifically designed to help geographers identify multiscale phenomenon, reflecting how the activity occurring at a local scale actually contribute to higher order dynamics.

Nodes (towns) are mapped to their geographical coordinates while edges (t_1, t_2) connecting two towns t_1 and t_2 consists in drawing them in 3D using line segments whose (x, y) coordinates of endpoints correspond to geographical coordinates of towns. Now, the dataset comes equipped with several edge attributes such as the distance between towns or the commuter traffic. Using one of these attributes, each segment is additionally assigned a z -coordinate. An orthographic projection is then used to obtain the final 2D image on the screen, where the projection plane is perpendicular to z axis (see Fig. 8).

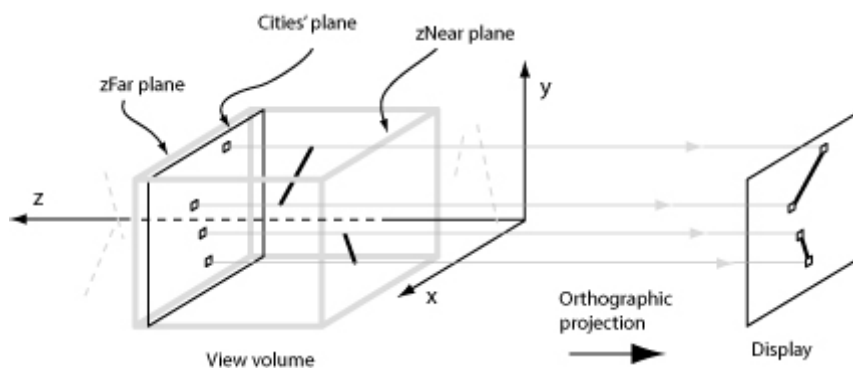


Figure 8. The commuter network as it forms over the French territory. Daily travels link cities. The underlying graph formed of cities (nodes) and daily travels (links) contains 33 thousand nodes and about 2 million links. Relevant interaction allows to filter links according to various metrics.

The work is ongoing. Variations on the participation coefficient [68] can be used in order to sort out edges relatively to a region or to any other cluster. We are now investigating different metrics and coloring schemes easing the visual comparison of edges going out of a same department, for instance, thus taking the total ow emerging from a same cluster and looking at the relative weight of an edge with respect to the cluster it goes out from.

7. Contracts and Grants with Industry

7.1. Strategic Watch

Participants: Guy Melançon, Maylis Delest, David Auber, Jean-Philippe Domenger, Patrick Mary, Faraz Zaidi.

- Project: FIVE Fouille Interactive, Veille, Visualisation et Exploration
- Call: ANR Software Technology (RNTL)
- start/end April 2007 – April 2009
- Budget: 964 970 euros (total) / 390 360 euros (grant) / 25 970 euros (INRIA GRAVITÉ)

We are involved in a project with two small size companies, AMI Software (<http://www.amisw.com>) developing text mining software and PIKKO Software (<http://www.pikko-software.com>) developing visualization software components. This project is funded through the national ANR RNTL FIVE (Software Technology 2007 call).

The project emerged from the need for companies to deal with large volume of new, inhomogeneous information on a daily basis with the aim of exploiting this information through their decision making process. This state of fact is not new: this is partly what risk management and strategic watch are concerned with. Collecting and analyzing this information has but a few things to share with information retrieval as it was classically designed. The surveillance and the collection of information is now performed on implicit domains and no standard methodology or tool has yet established as a standard.

The FIVE project aims at developing visual approaches to support information collection through graphical representations and interaction well suited to tasks conducted by information analysts.

AMI Software (<http://www.amisw.com>) develops software tools to enhance decision making. AMI's technology supports text mining, entity and concept identification in open and non structured information. PIKKO develops visualization components that can be easily integrated into existing information systems.

We bring in our expertise in data (graphs) analysis and graph visualization and navigation. The project requires that we conduct a fine-grained study of potential visualization in the context of text analysis and visualization.

8. Other Grants and Activities

8.1. ANR Systryp

Participants: Guy Melançon, David Auber, 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 modelling of biochemical networks relates to a lack of general compatibility between static and dynamic modelling. 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 modelling, to a detailed representation derived from dynamic modelling. 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 modelling 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 trypanosomes 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 modelling. Predictions based on the modelling 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 modelling 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 modelling 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. ANR SPANGEO

Participants: Guy Melançon, David Auber, Romain Bourqui, Maylis Delest, Trung Tien Phan Quang, Pierre-Yves Koenig.

- Project: SPANGEO (Multi-level modeling of Spatial Networks in Geography)
- Call: ANR Massive Data
- start/end January 2006 – January 2009
- Budget: 2 420 000 euros (grant all partners)

Although it started even before the team was created, we mention this project as part of our activities because all members are involved in its development. The project gave us the opportunity to develop deep and long-lasting relationships with colleagues working in geography.

Geographers studying transportation and communication systems, economical and social networks or urban systems, are confronted with matrices of increasing size that are more and more difficult to process and represent. The increased complexity of these networks, on various space scales, makes it hard to deal with the entirety of the networks. Their multilevel nature is thus hard to identify leaving the analyst with an incomplete comprehension of the systems. The project aims at designing and developing operational methods for modeling and representing networks as multilevel spaces. A first collaboration looking at the multilevel representation of the air passenger traffic worldwide network, had brought significant results, helping to model and visualize spatial networks [1]. However, many aspects had to be carefully re-examined and developed before extending the methodology and tools to other issues in spatial geography.

The project thus requires the extension of graph multi-level clustering and to various geographical problems. In particular, extensions to weighted or directed graphs, or hypergraphs must be studied. Moreover, graphical conventions were for a long time developed in cartography. It is mandatory to see how those conventions can be taken into account when representing spatial networks because of their strong semantics (proportional size of the nodes, of the edges, distances between nodes, colors of nodes and colors of edges, for instance). The cooperation between geographers and computer scientists aims at producing methods, algorithms and tools able to deal with large geographical matrices and to implement a set of approaches for modeling and representing these graphs. It is necessary to work collectively at the development of methods and software tools offering a multilevel treatment of the graphs dedicated to the fields of geography and related disciplines. The dissemination of these methods and tools in the geographical scientific community will allow a better operational treatment of spatial networks and a better definition of social and territorial consequences of their evolutions.

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.

9. Dissemination

9.1. Program committees and related activities

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).

Our team is organizing a one-day tutorial session on advanced Tulip programming during the EGC 2009 next January.

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

- EGC (French) Yearly conference on Data Mining and Knowledge Management (David Auber, since 2007)
- SYNASC Yearly Conference, Timisoara, Romania (G. Melançon, since 2005)
- CGIV Computer Graphics, Imaging, Visualization (G. Melançon, since 2005)
- ACM Symposium on Software Visualization SoftVis 2006 and SoftVis 2008 (David Auber)
- Eurographics IEEE EuroVis Symposium on Visualization (D. Archambault 2008, D. Auber since 2007, G. Melançon 2008)
- IEEE Information Visualization International Conference (Guy Melançon, since 2001)

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)
- Jury member for the ANR Massive Data program 2008 – 2011 (Guy Melançon)
- NSF Merit Review, see <http://www.nsf.gov/> (G. Melançon, June 2008)
- Regional call Digiteo – closely related to the “Pôle d'excellence System@tic Ile-de-France” (G. Melançon, April 2008)

10. Bibliography

Major publications by the team in recent years

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

- [2] D. ARCHAMBAULT, T. MUNZNER, D. AUBER. *Smashing Peacocks Further: Drawing Quasi-Trees from Biconnected components*, in "IEEE Transactions on Visualization and Computer Graphics (Proc. Vis/InfoVis 2006)", vol. 12, n^o 5, September-October 2006, p. 813–820.
- [3] D. ARCHAMBAULT, T. MUNZNER, D. AUBER. *Grouse: Feature-Based, Steerable Graph Hierarchy Exploration*, in "Eurographics/ IEEE-VGTC Symposium on Visualization", K. MÜSETH, T. MÖLLER, A. YNNERMAN (editors), Eurographics Association, Eurographics, 2007, p. 67–74.
- [4] D. AUBER. *Tulip - A huge graph visualization framework*, P. MUTZEL, M. JÜNGER (editors), Mathematics and Visualization Series, Springer Verlag, 2003.
- [5] 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.
- [6] 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", vol. 10, n^o 2, 2006, p. 329-351.
- [7] R. BOURQUI, D. AUBER, V. LACROIX, F. JOURDAN. *Metabolic network visualization using constraint planar graph drawing Algorithm*, in "IV 06: Proceedings of the 10 International Conference on Information Visualisation (IV06)", IEEE Computer Society, 2006.
- [8] R. BOURQUI, L. COTTRET, V. LACROIX, D. AUBER, P. MARY, M.-F. SAGOT, F. JOURDAN. *Metabolic network visualization eliminating node redundancy and preserving metabolic pathways*, in "BMC Systems Biology", vol. 1, n^o 29, 2007.
- [9] F. CHEVALIER, M. DELEST, J.-P. DOMENGER. *A heuristic for the retrieval of objects in video in the framework of the rough indexing paradigm*, in "Signal Processing : Image Communication (SPIC)", vol. 22, 2007, p. 622–634.
- [10] M. DELEST, A. DON, J. BENOIS-PINEAU. *DAG based visual interfaces for navigation in index vidéo content*, in "Journal Multimedia Tools and Applications", vol. 31, n^o 1, Oct 2006, p. 51-72.
- [11] 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.
- [12] M. DELEST, T. MUNZNER, D. AUBER, J.-P. DOMENGER. *Exploring InfoVis Publication History with Tulip (2nd place - InfoVis Contest)*, in "IEEE Symposium on Information Visualization", IEEE Computer Society, 2004, 110.
- [13] H. ROMAN, M. ALBERGANTE, M. COLOMBO, F. CROCCOLO, F. MARINI, C. RICCARDI. *Modeling cross correlations within a many-assets market.*, in "Physical Review E Statistical, Nonlinear, and Soft Matter Physics", 2006.

Year Publications

Articles in International Peer-Reviewed Journal

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

[15] 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/>.

Articles in National Peer-Reviewed Journal

[16] P.-Y. KOENIG, G. MELANÇON. *Dagmap : exploration interactive de relations d'héritage*, in "Revue d'Intelligence Artificielle", vol. 22, 2008-05, p. 355-370, <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272829/en/>.

International Peer-Reviewed Conference/Proceedings

[17] O. BARRIERE, E. LUTTON, C. BAUDRIT, M. SICARD, B. PINAUD. *Modeling human expertise on a cheese ripening industrial process using GP*, in "10th international conference on Parallel Problem Solving from Nature: PPSN X", Lecture Notes in Computer Science, vol. 5199, 2008, p. 859–868, <http://hal.inria.fr/inria-00335956/en/>.

[18] G. MELANÇON, A. SALLABERRY. *Edge Metrics for Visual Graph Analytics: A Comparative Study*, in "12th International Conference on Information Visualisation, London, UK", IEEE Computer Society, 2008-07, p. 610-615, <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272784/en/>.

[19] P. SIMONETTO, D. AUBER. *Visualise Undrawable Euler Diagrams*, in "12th International Conference on Information Visualisation, London, UK", 2008-07-25, p. 594-599, <http://hal.archives-ouvertes.fr/hal-00319119/en/>.

National Peer-Reviewed Conference/Proceedings

[20] F. CHEVALIER, M. DELEST, J.-P. DOMENGER. *Recherche de motifs quasi-similaires dans des graphes*, in "EGC 2008, France", 2008, –, <http://hal.archives-ouvertes.fr/hal-00306972/en/>.

[21] P.-Y. KOENIG, G. MELANÇON. *DagMap : exploration interactive de relations d'héritage*, in "EGC'08: Extraction et Gestion des Connaissances, Sophia Antipolis, Nice, France", vol. 001, 2008, p. 001-015, <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00276097/en/>.

Workshops without Proceedings

[22] R. BOURQUI, F. ZAIDI, F. GILBERT, U. SHARAN, P. SIMONETTO. *VAST 2008 Challenge: Social network dynamics using cell phone call patterns*, in "IEEE VisWeek Conference Compendium", 2008, p. 138-139, <http://hal.archives-ouvertes.fr/hal-00335647/en/>.

[23] 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", 2008-05, <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272786/en/>.

[24] G. MELANÇON. *Visualisation: Cartographie ? Interaction ? Travail collaboratif ?*, in "Colloque Carto 2.0", C. ASCHENBRENNER, C. TRICOT (editors), 2008.

- [25] B. PINAUD, C. BAUDRIT, M. SICARD, P.-H. WUILLEMIN, N. PERROT. *Validation et enrichissement interactifs d'un apprentissage automatique des paramètres d'un réseau bayésien dynamique appliqué aux procédés alimentaires*, in "Journées Francophone sur les Réseaux Bayésiens, Lyon France", 2008, <http://hal.archives-ouvertes.fr/hal-00259891/en/>.
- [26] 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.
- [27] U. SHARAN, G. MELANÇON. *VAST 2008 Challenge: Social Collaboration and Faction Discovery Using Disagreement Graphs*, in "IEEE VisWeek Conference Compendium", 2008, p. 154-155, <http://hal.archives-ouvertes.fr/hal-00335647/en/>.

Scientific Books (or Scientific Book chapters)

- [28] 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, vol. 4950, Springer, 2008, p. 154-175, <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272779/en/>.
- [29] D. AUBER, G. MELANÇON, R. BOURQUI. *Mining Networks through Visual Analytics: Incremental Hypothesis Building and Validation*, in "Mining Massive Data Sets for Security", F. F. SOULIÉ, D. PEROTTA, J. PISKORSKI, R. STEINBERGER (editors), IOS Press, 2008, p. 204-211, <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00272781/en/>.
- [30] G. MELANÇON. *Réseaux sociaux, espaces collaboratifs et visualisation d'information*, in "Information & visualisation - enjeux, recherches et applications", S. CHAUVIN (editor), Cepadues, 2008, -, http://www.cephadues.com/livre_details.asp?l=856.

References in notes

- [31] S. BORNHOLDT, G. SCHUSTER (editors). *Handbook of Graphs and Networks: From the Genome to the Internet*, Wiley-VCH, 2003.
- [32] U. BRANDES, T. ERLEBACH (editors). *Network Analysis*, Lecture Notes in Computer Science, vol. 3418, Springer, 2005.
- [33] C. JOHNSON, R. MOORHEAD, T. MUNZNER, H. PFISTER, P. RHEINGANS, T. S. YOO (editors). *NIH-NSF Visualization Research Challenges Report*, IEEE Computer Society, 2006.
- [34] J. J. THOMAS, K. A. COOK (editors). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE Computer Society, 2006.
- [35] J. ABELLO, F. VAN HAM, N. KRISHNAN. *ASK-GraphView: A Large Scale Graph Visualization System*, in "IEEE Trans. on Visualization and Computer Graphics (Proc. Vis/InfoVis '06)", vol. 12, n^o 5, 2006, p. 669-676.
- [36] I. H. ANSOFF. *Managing strategic surprise by response to weak signals*, in "California Management Review", vol. XVII, n^o 2, 1976, p. 21-33.

-
- [37] 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, vol. 2265, Springer-Verlag, 2001, p. 335-337.
- [38] D. AUBER. *Using Strahler Numbers for Real-Time Visual Exploration of Huge Graphs*, in "International Conference on Computer Vision and Graphics", Journal of WSCG, vol. 10/1-3, 2002, p. 56-69.
- [39] D. AUBER, M. DELEST. *A Clustering Algorithm for Huge Trees*, in "Advances in Applied Mathematics", vol. 31, n° 1, 2003, p. 46-60.
- [40] 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 Visualization", IEEE Computer Society, 2003, p. 124-126.
- [41] D. AUBER, P. MARY. *Mise en place d'un mécanisme de plugins en C++*, in "Programmation sous linux", vol. 5, n° 5, 2006, p. 74-79.
- [42] G. BAILLY, D. AUBER, L. NIGAY. *2m: un espace de conception pour l'interaction bi-manuelle*, in "UbiMob '05: 2nd French-speaking conference on Mobility and ubiquity computing", ACM Press, 2005, p. 177-184.
- [43] 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, to appear.
- [44] G. BAILLY, D. AUBER, L. NIGAY. *Navrna: Visualization - exploration - edition of RNA*, in "International Conference on Advanced Visual Interfaces (AVI '06)", ACM, 2006, to appear.
- [45] M. Q. W. BALDONADO, A. WOODRUFF, A. KUCHINSKY. *Guidelines for Using Multiple Views in Information Visualization*, in "Advanced Visual Interfaces", 2000, p. 110-119.
- [46] 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, vol. 2265, 2002, 477.
- [47] A. BEYGELZIMER, S. KAKADE, J. LANGFORD. *Cover trees for nearest neighbor*, in "23rd international conference on Machine learning", vol. 148, ACM Press, 2006, p. 97-104.
- [48] J. BINDER, G. HAAG, G. RABINO. *Analysis and modelling of commuter flows: application to the regions of Stuttgart and Turin*, in "Jahrbuch der Regionalwissenschaften", vol. 23, n° 117, 2003, p. 117-139.
- [49] 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.
- [50] 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", vol. 71, n° 7-9, 2008, 12571273.

-
- [51] 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.
- [52] F. CHEVALIER, D. AUBER, A. TELEA. *Structural Analysis and Visualization of C++ Code Evolution using Syntax Trees*, in "International Workshop on Principles of Software Evolution (IWPSE)", 2007, p. 90-97.
- [53] 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.
- [54] E. 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.
- [55] 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.
- [56] S. N. DOROGOVTSSEV, J. F. F. MENDES. *Evolution of Networks : From Biological Nets to the Internet and WWW*, Oxford University Press, 2003.
- [57] P. EADES, M. L. HUANG. *Navigating Clustered Graphs using Force-Directed Methods*, vol. 1547, 1998, p. 374–383.
- [58] J. ELLSON, E. 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, vol. 2265, 2002, 483.
- [59] J. ELLSON, E. GANSNER, L. KOUTSOFIOS, S. NORTH, G. WOODHULL. *Graphviz and Dynagraph - Static and Dynamic Graph Drawing Tools*, in "Graph Drawing Software", M. JUNGER, P. MUTZEL (editors), Springer-Verlag, 2003, p. 127-148.
- [60] J.-D. FEKETE. *The InfoVis Toolkit*, in "10th IEEE Symposium on Information Visualization, Austin, TX", IEEE Press, 2004, p. 167-174.
- [61] 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", 1077-2626, vol. 9, n^o 3, 2003, p. 378-394.
- [62] L. C. FREEMAN. *The Development of Social Network Analysis: A Study in the Sociology of Science*, Empirical Press, Vancouver, 2004.
- [63] L. C. FREEMAN. *A set of measures of centrality based upon betweenness*, in "Sociometry", vol. 40, 1977, 3541.
- [64] G. W. FURNAS. *Generalized Fisheye Views*, in "Human Factors in Computing Systems CHI '86", ACM Press, 1986, p. 16-23.

-
- [65] G. GAINANT, D. AUBER. *Arna: Interactive comparison and alignment of rna secondary structure*, in "IEEE Symposium on Information Visualisation", IEEE Computer Society, 2004, 8.
- [66] G. GAINANT, D. AUBER. *Arna: Interactive comparison and alignment of rna secondary structure*, in "IEEE Symposium on Information Visualisation", IEEE Computer Society, 2004, 8.
- [67] E. R. GANSNER, S. C. NORTH. *An open graph visualization system and its applications to software engineering*, in "Software Practice and Experience", vol. 30, n^o 11, 2000, p. 1203-1233.
- [68] R. GUIMERÀ, L. A. N. AMARAL. *Cartography of complex networks: modules and universal roles*, in "Journal of Statistical Mechanics: Theory and Experiment", vol. February, 2005.
- [69] 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, vol. 3202, Springer, 2004, p. 18-26.
- [70] D. J. HAND. *Statistics and data mining: intersecting disciplines*, in "ACM SIGKDD Explorations Newsletter", vol. 1, n^o 1, 1999, p. 16-19.
- [71] F. HARARY. *Graph Theory*, Addison-Wesley, Reading, MA, USA, 1972.
- [72] 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.
- [73] I. HERMAN, M. S. MARSHALL, G. MELANÇON. *Graph Visualisation and Navigation in Information Visualisation: A Survey*, in "IEEE Transactions on Visualization and Computer Graphics", vol. 6, n^o 1, 2000, p. 24-43.
- [74] I. HERMAN, G. MELANÇON, M. DELEST. *Tree Visualisation and Navigation Clues for Information Visualisation*, in "Computer Graphics Forum", vol. 17, n^o 2, 1998, 153165.
- [75] INRIA. *Plan stratégique 2003-2007*, 2003.
- [76] F. IRAGNE, M. NIKOLSKI, B. MATHIEU, D. AUBER, D. SHERMAN. *Proviz: protein interaction visualization and exploration*, in "Bioinformatics", vol. 21, n^o 2, 2005, p. 272-274.
- [77] 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.
- [78] D. KEIM, H.-P. KRIEGEL. *Visualization Techniques for Mining Large Databases: A Comparison*, in "IEEE Transactions on Knowledge and Data Engineering", vol. 8, n^o 6, 1996, p. 923-936.
- [79] 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.

- [80] P. KUNTZ, B. PINAUD, R. LEHN. *Minimizing crossings in hierarchical digraphs with a hybridized genetic algorithm*, in "Journal of Heuristics", vol. 12, n^o 1, 2006, p. 23-36.
- [81] G. LOHSE, K. BIOLSI, N. WALKER, H. RUETER. *A classification of visual representations*, in "Communications of the ACM", vol. 37, n^o 12, 1994, p. 36-49.
- [82] 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", vol. 11, n^o 6, 2001, p. 703-715.
- [83] G. MELANÇON, Y. CHIRICOTA. *Visually mining relational data*, in "International Review on Computers and Software", 2007, p. 1-14.
- [84] G. MELANÇON, P.-Y. KOENIG, C. BOHAN, B. GAUTHIER. *Combining DagMaps and Sugiyama Layout for the Navigation of Hierarchical Data*, in "IV'07: 11th International Conference on Information Visualisation, Zurich, Switzerland", IEEE Computer Society, 2007.
- [85] T. PATTISON, R. VERNIK, M. PHILLIPS. *Information Visualization using Composable Layouts and Visual Sets*, in "Proc. of the 2001 Asia-Pacific Symposium on Information Visualization", 2001, p. 1-10.
- [86] R. PATUELLI, A. REGGIANI, S. GORMAN, P. NIJKAMP, F. BADE. *Network analysis of commuting flows: a comparative static approach to german data*, in "Networks and Spatial Economics", vol. 7, n^o 4, 2007, p. 315-331.
- [87] 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.
- [88] L. PETER, H. R. VARIAN. *How Much Information*, n^o March 2006, 2003.
- [89] 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.
- [90] 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.
- [91] T. SIKORA. *The MPEG-7 Visual Standard for Content Description An Overview*, in "IEEE Transactions on Circuits and Systems for Video Technology", vol. 11, n^o 6, 2001, p. 696-702.
- [92] P. SIMONETTO, D. AUBER. *A Heuristic for the Construction of Intersection Graphs*, Technical report, LaBRI, University of Bordeaux, 2008.
- [93] M. TORY, T. MÖLLER. *Rethinking Visualization: A High-Level Taxonomy*, in "IEEE Symposium on Information Visualization 2004", 2004, p. 151-158.
- [94] M. O. WARD. *A taxonomy of glyph placement strategies for multidimensional data visualization*, in "Information Visualization (Palgrave)", vol. 1, n^o 3/4, 2002, p. 194-210.

-
- [95] C. WARE. *Information Visualization: Perception for Design*, Morgan Kaufmann Publishers, Orlando, FL, 2000.
- [96] D. WATTS, S. H. STROGATZ. *Collective dynamics of "small-world" networks*, in "Nature", vol. 393, 1998, p. 440-442.
- [97] D. J. WATTS. *Six Degrees: The Science of a Connected Age*, 1st, W. W. Norton & Company, 2003.
- [98] D. J. WATTS. *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, 1999.
- [99] W. DE NOOY, A. MRVAR, V. BATAGELJ. *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, 2004.
- [100] S. DOS SANTOS, K. BRODLIE. *Gaining understanding of multivariate and multidimensional data through visualization*, in "Computers & Graphics", vol. 28, n^o 3, 2004, p. 311-325.
- [101] J. J. VAN WIJK, F. VAN HAM. *Interactive Visualization of Small World Graphs*, in "IEEE Symposium on Information Visualisation, Austin, TX, USA", T. MUNZNER, M. WARD (editors), IEEE Computer Science press, 2004.
- [102] J. VAN WIJK. *The Value of Visualization*, in "IEEE Visualization", C. SILVA, E. GROELLER, H. RUSHMEIER (editors), IEEE Computer Society, 2005, p. 79-86.