Link Analysis and Visualization

01.07. – 06.07.2001

organized by

Ulrik Brandes (Universität Konstanz)
David Krackhardt (Carnegie Mellon University)
Roberto Tamassia (Brown University)
Dorothea Wagner (Universität Konstanz)

The purpose of this seminar was to introduce to each other researchers working on different aspects and applications of link analysis and visualization in order to strengthen the algorithmic foundations of this rapidly emerging, highly interdisciplinary, field.
Link analysis explores associations among entities of arbitrary type. It is increasingly recognized as a fruitful extension of categorical approaches to data analysis in a fast growing number of application domains. Example applications are the analysis of linkages on the Web (search engines, site maps), network traffic monitoring (Web caching, public transport), data mining (e-commerce, telecommunications services), social network analysis (social structures, policy making), text analysis (coreference, cocitation), decision support (financial markets, logistics), or fraud detection (money laundring, calling cards).
Typical objectives in these applications are the identification of central or bottleneck entities, structural patterns and trends, effective modifications, hidden or missing data, substructures, appropriate levels of aggregation, similarities among data sets, etc., and visualization has proven crucial in assisting humans to comprehend complex relational structures and identify unexpected patterns.

Participants

There were 24 invited participants from 5 countries on 3 continents, many of which have never met before. Since a substantial number of them also have not been to Dagstuhl before, they freshly experienced amazement about excellent working atmosphere and facilities known all to well to those that like to return for exactly that reason. On a related note, there have been several reports about weight problems following a week of Dagstuhl catering.
Despite some unforeseen short- and without-notice cancellations, the participants collectively covered unusually many areas of research, with backgrounds ranging from algorithmic graph theory, algorithms & data structures (including very large graphs), and graph drawing, via knowledge discovery and visualization to organizational theory and social network analysis.

Vladimir Batagelj, Univerza v Ljubljani
James Blythe, University of Southern California
Ulrik Brandes, Universität Konstanz
Steven R. Corman, Arizona State University
Sabine Cornelsen, Universität Konstanz
Klaus Holzapfel, Technische Universität München
David Jensen, University of Massachusetts
Jeffrey C. Johnson, East Carolina University
Michael Kaufmann, Universität Tübingen
Stephen Kobourov, University of Arizona
David Krackhardt, University of California
Lothar Krempel, Max Planck Institut für Gesellschaftsforschung, Köln
Giuseppe Liotta, Universitá di Perugia
Alberto Marchetti-Spaccamela, Universitá di Roma “La Sapienza”
Cathleen McGrath, Loyola Marymount University
Margaret Mitchell, Macquarie University
Rolf H. Möhring, Technische Universität Berlin
Stefan Näher, Universität Trier
Marcus Raitner, Universität Passau
Marc Rittberger, Universität Konstanz
Maja Ruby, Universität Kaiserslautern
Roberto Tamassia, Brown University
Ioannis G. Tollis, University of Texas at Dallas
Dorothea Wagner, Universität Konstanz

Program

The scientific program served to introduce general areas of research related to the theme of the seminar as well as personal research backgrounds. Each morning was opened with a longer talk giving an overview of a hot topic. Several presentations will appear in refereed conference proceedings only later this year, and some others contained work in progress. Two extra sessions were included in the program: The ever-favorite open problem session, and a special session on current efforts within the graph drawing research community to define an XML standard for graphs and graph drawings.
The main purpose of this seminar, however, was to stimulate the exchange of ideas between researchers interested in link analysis and visualization for various reasons. To this end, the concept of "PTH" was imported from early editions of the Annual Sunbelt Social Network Conference: The Sunbelt conference series started out in 1981 and, as the name implies, was initially held in southern states of the U.S. which are collectively known as the Sunbelt. Being social networkers, the organizers felt strongly about the importance of personal relations, and therefore did not schedule any presentations for the time between noon and 4pm. Because of generally sunny conditions at conference locations, those time slots became known as "peak tanning hours," or PTH for short. Schloß Dagstuhl proved to be another perfect environment for such peer-to-peer and small-group collaborations.

Monday, July 2nd

9:00 Welcome
9:15 Stephen Kobourov:
Drawing Graphs with Fat Edges
9:45 Ulrik Brandes:
On Formalizing Structural Importance in Networks

11:00 Jeff Johnson:
Exploring Networks Interactively in 3D Space
11:30 Lothar Krempel:
Network and Information Visualization: Patterns in Economic Networks

16:00 Marcus Raitner:
Graphs with a Hierarchy – Models and Operations
16:30 Sabine Cornelsen:
Visualizing Small Cuts
17:00 Margaret Mitchell:
Use of Node-Link Graph Analysis
in Generating Instructions for Multiple Users

Tuesday, July 3rd

9:00 David Jensen:
Learning Statistical Models from Relational Data

10:45 Cathleeen McGrath:
User Studies Testing the Effectiveness of Visualization Techniques for Communicating Social Network Information

11:15 Ioannis G. Tollis:
A Framework for the Visualization of Networks

16:30 Marc Rittberger:
Quality Evaluation of Search Engines

17:00 Open Problem Session (chaired by Giuseppe Liotta)

Wednesday, July 4th

9:00 Vladimir Batagelj:
Analysis and Visualization of Large Networks with Program Pajek

10:30 Forum on Graph (Drawing) Exchange Formats
Roberto Tamassia: *Background on XML*
Ulrik Brandes: *Current State of the GraphML Proposal Discussion*

14:00 Exkursion
Hiking

Thursday, July 5th

9:00 Steven R. Corman:
Centering Resonance Analysis of Text and Conversation, with a Special Application to Analyzing Group Discourse

10:45 James Blythe:
Highly Customizable Tools for Graph Drawing

11:15 Michael Kaufmann:
On the Visualization of Web Site Maps

16:30 David Krackhardt interpreting Valdis Krebs:
Corporate Networks

17:00 Dorothea Wagner:
Social Network Analysis and Visualization

Friday, July 6th

9:00 Klaus Holzapfel:
On the Small World Phenomenon

10:00 Closing Remarks

Abstracts

Analysis and Visualization of Large Networks Using Pajek

Vladimir Batagelj (University of Ljubljana)

The first part of the contribution is a short overview of program Pajek. Pajek (Slovene word for Spider) is a program, for Windows (32 bit), for analysis and visualization of large networks. It is freely available, for noncommercial use, at its homepage:

`http://vlado.fmf.uni-lj.si/pub/networks/pajek/`

I am developing Pajek together with Andrej Mrvar. Some procedures were contributed also by Matjaž Zaveršnik.
The main goals in the design of Pajek are:

- to support abstraction by (recursive) factorization of a large network into several smaller networks that can be treated further using more sophisticated methods;
- to provide the user with some powerful visualization tools;
- to implement a selection of efficient (subquadratic) algorithms for analysis of large networks.

Besides ordinary networks, that can be imported in several formats (Pajek's native, Ucinet DL, GEDCOM, ball and stick, MDL, Vega), Pajek supports also 2-mode networks and temporal networks.
Pajek is based on six data structures: network, partition, cluster, permutation, hierarchy and vector. It is essentially a collection of routines operating on these data structures. Typical sequences of applications of routines can be 'encapsulated' as macros.
We developed efficient algorithms for determining *main paths* in acyclic networks, *cores* and counting *triads*.

Special emphasis is given to automatic generation of network layouts. Several standard algorithms for automatic graph drawing are implemented: spring embedders, layouts determined by eigenvectors, drawing in layers (genealogies and other acyclic structures), fish-eye views and block (matrix) representation.
These algorithms were modified and extended to enable additional options: drawing with constraints (optimization of the selected part of the network, fixing some vertices to predefined positions, using values of lines as similarities or dissimilarities), drawing in space. Pajek also provides tools for manual graph editing.
Pajek supports several output graphic formats where layouts can be examined by special 2D and 3D viewers or plug-ins: Encapsulated PostScript (EPS) – GSView; VRML (WRL) – CosmoPlayer; MDLMOL (MOL) – Chime; Kinemages (KIN) – Mage; Scalable Vector Graphics (SVG) – SVGview.
In the second part 2-mode network analysis and different approaches to analysis and visualization of large networks, implemented in Pajek:

- global properties of network;
- distributions (degree, triads, cores, values, components,...);
- connectivity (weak/strong components, biconnected components);
- clustering (visual: eigenvectors, spring-embedders; methods from data analysis, blockmodeling, special methods); shrinking, hierarchy;
- neighborhoods of important vertices (selected vertex, indices, input values, ...);
- special substructures (cores, main paths, ...);
- pattern searching.

are presented.
The HTML version of the contribution is available at

`http://vlado.fmf.uni-lj.si/pub/conf/dagstuhl.01/`

User-Customizable Tools for Graph Visualization

James Blythe (University of Southern California)

Social networks are created and explored in an extremely wide variety of situations, and programs for their visualization and exploration need to be very

flexible. Open Krackplot is a visualization program that allows the user to view and modify the algorithms used for both network layout and presentation, so one can customize the drawing in ways that would otherwise require re-programming. We make use of recent work in knowledge acquisition and ontologies to structure Open Krackplot so that it is relatively easy to modify, and use an editor based on automatically generated structured english paraphrases to shield the user from programming constructs. A rich visualization taken from Krempel and Plümper '99 can be approximately re-constructed in a few minutes using the tool. It is therefore a promising platform for rapidly testing and exploring graph presentations.

On Formalizing Structural Importance in Networks

Ulrik Brandes (University of Konstanz)

Structural importance in a network is used as an indication to argue about questions such as "Who is the most important person in a group of people?" or "Which procedure is crucial in the design of a control flow?"
The determination of structurally important vertices therefore is one of the most important uses of graph theory in network analysis. However, the incredible variety of existing measures is only loosely connected and lacks a formal framework allowing classification of measures or general statements about families of measures. We present first steps in this direction by introducing a framework to normalize measures on any graph to a probability distribution and by proposing a minimal axiomatization based on the implicit intuition behind popular importance measures.

Centering Resonance Analysis of Text and Conversation, with a Special Application to Analyzing Group Discourse

Steven R. Corman (Arizona State University)

Traditional computer text analysis uses an inverse document frequency scheme to determine the importance of words. This is an information theoretic approach that presents theoretical, operational, and practical scheme to determine the

importance of words. This is an information theoretic approach that presents theoretical, operational, and practical problems. Centering Resonance Analysis (CRA) is a new, discourse theoretic approach that employs linguistic analysis to render a text as a set of noun phrases. These phrases contain centering tokens, words authors and speakers employ to maintain coherent communication. Linking these words into a network and analyzing their betweenness centrality yields an abstract representation of the text's important words and their interrelations. CRA networks support a variety of applications including visualization of document of document collections, full text searching, spatial analysis and modeling, and dynamic analysis of conversation. These applications are illustrated, and a special analysis is presented of the U.S. Supreme Court's deliberations in the Bush vs. Gore case concerning the 2000 Presidential Election.

Visualizing Small Cuts

Sabine Cornelsen (Universität Konstanz)

There exist basically two graph models that represent small edge-cuts of undirected connected multi-graphs: The cactus model of all minimum cuts developed by Dinitz, Karzanov, and Lomonosov and the 2-level cactus for all minimum and minimum+1 cuts of Dinitz and Nutov which generalizes a model of Galil and Italiano for the case of connectivity one and a model of Dinitz and Westbrook for the case of connectivity two.
The contents of the talk is divided into two parts. First, we show how to utilize the cactus model of all minimum cuts of a planar graph G to visualize the minimum cuts of G in a planar drawing in such a way that the two vertex subsets of every minimum cut are separated by a simple closed curve and vice versa. This goal is achieved by extending techniques for drawing hierarchically clustered graphs.
Second, we study planarity of the 2-level cactus model. We give a sufficient planarity criterion in terms of projection paths over a spanning subtree of a graph. Using this criterion, we show that the 2-level cactus of a graph G – extended by some auxiliary edges in case of odd connectivity – is planar if the cardinality of a minimum edge-cut of G is not equal to 2, 3 or 5. On the other hand, we give examples for non-planar 2-level cacti of graphs with these connectivities.
This is joint work with Ulrik Brandes, Yefim Dinitz, Christian Fieß, and Dorothea Wagner.

The Small World Phenomenon

Klaus Holzapfel (Technische Universität München)

Within the last decades it was observed that in social networks the distance between two arbitrary people seems to be far less than one would expect. Further investigations lead to a definition of the so called 'small world' phenomenon, i.e. the property of graphs which model those networks to be dense, have small diameter and the tendency to cluster.
In the last few years several models have been set up to describe such graphs. Further there was observed that many large real world networks, such as the Internet or the Hollywood graph, have a degree distribution corresponding to the power-law, i.e. the number of vertices with degree d is proportional to $d^{-}r$ for fixed r. Based on these results generators have been built to provide a possibility to simulate such networks.
In this talk we present some popular and promising models and generators for small world graphs.

Learning Statistical Models from Relational Data

David Jensen (University of Massachusetts)

A large number of techniques exist to visualize relational data. However, relatively few techniques exist to discover statistical regularities in relational data and use those regularities to construct predictive models. We are studying a number of techniques to construct statistical models, and we are applying those models to several large data sets. These include predicting attributes of companies based on corporate relationships in the banking and chemical industries, predicting fraud based on calling patterns in wireless telephone networks, and predicting attributes of movies based on relationships in the entertainment industry. In addition, we are developing tools for querying, sampling, and transforming large sets of relational data.

Exploring Networks Interactively in 3D Space

Jeffrey C. Johnson (East Carolina University)

Kinemages are designed to enable understanding of the internal structure of a 3D object or the multi-dimensional relationships among a set of objects. Visual representations are used to enhance the viewing of 3D connections and contacts as well as the ability to be moved interactively in real time. Any point can be identified simply by clicking on it, and any distance measured in the metric of the displayed objects. Recent additions to MAGE for network visualization include arrows to show direction of a relationship, rings to encircle nodes so that multiple connectivity at an intersection can be seen rather than be buried inside a ball, and cross-picking between cells of tabular data and nodes in the 3D graphics presentation (e.g., superimposing non-network information onto a 3D structural representation). This presentation provides examples of applications of Mage for exploring network structure in both the social and ecological sciences. Particular attention is paid to exploring theoretically important features of both nodes and arcs as they relate to any given structure and means for the animation of network dynamics.
This is joint work with David Richardson and Jane Richardson.

On the Visualization of Web Site Maps

Michael Kaufmann (University of Tübingen)

The World-Wide Web has created an extremely huge but 'messy' information space which is hard to overview. Site maps as alternative views of Web sites have been proposed to assist the user in navigating the hyperspace. As Web localities are subject to frequent change and redesign, it is especially important to provide a system for automatic generation of such site maps from various data sources instead of creating them manually. In the talk, we presented two new structure-based and user-based approaches. We elaborated the necessary concepts of data extractions and visualizations. Finally, we demonstrated the results of our implementation using the yFiles by various screen shots.

Drawing Graphs with Fat Edges

Stephen Kobourov (Arizona State University)

We introduce the problem of drawing with "fat" edges. Traditionally, graph drawing algorithms represent vertices as circles and edges as closed curves connecting the vertices. Here we consider the problem of drawing graphs with edges of variable thickness. The thickness of an edge is often used as a visualization cue, to indicate importance, or to convey some additional information. We present a model for drawing with fat edges and a corresponding polynomial time algorithm that uses the model.
We focus on a restricted class of graphs that occur in VLSI wire routing and show how to extend the algorithm to general planar graphs. We show how to take an arbitrary wire routing and convert it into a homotopic equivalent routing such that the distance between any two wires is maximized. Moreover, the routing uses the minimum length wires. Maximizing the distance between wires is equivalent to finding the drawing in which the edges are drawn as thick as possible. To the best of our knowledge this is the first algorithm that finds the maximal distance between any two wires and allows for wires of variable thickness. The previous best known result for the corresponding decision problem with unit wire thickness is the algorithm of Gao *et al.*, which runs in $O(kn^2 \log(kn))$ time and uses $O(kn^2)$ space, where n is the number of wires and k is the maximum of the input and output complexities. The running time of our algorithm is $O(kn + n^3)$ and the space required is $O(k+n)$. The algorithm generalizes naturally to general planar graphs as well.

David Krackhardt's Interpretation of Valdis Krebs' Consulting Using Network Graphs

David Krackhardt (Carnegie Mellon University)

Valdis Krebs is one of the most successful and prominent organizational consultants applying network analysis to business organizations today. This presentation of his work centered on three examples from his experience in the field. First, his graphs show how managers prefer to think in terms of formal organizational units (departments) and to draw pictures of the informal organization using circles representing these units. With a spring embedder, however, these pictures are much clearer and more informative, even to these same managers.

With Valdis' help, IBM studied a pharmaceutical firm, using four relational questions. They discovered that the firm had many holes in its network and concluded that efforts should be made to plug those holes. Also, they could see that one department was completely self-absorbed and that they needed to make bridges to the rest of the organization to generate new ideas.
The second example followed a merger by two firms. Network pictures were used to monitor the progress of integration of this merger. It was clear from the pictures that top management was not integrating very well – most of management was associating only with people in their original company.
Third, he provided an example of an inter-firm analysis, showing how the high tech industry is forming strategic alliances with many different kinds of firms in an attempt to expand and ensure profitability in their businesses.
Finally, Krackhardt provided an example of how the dynamics of a firm's response to a unionization attempt could be understood by looking at the social network ties in the firm. One individual was chosen to lead the group, but he was ineffective because he was completely isolated from the group in the friendship network. Second, a key player (Chris), who could have been very influential in getting the people to support the union, was over-embedded in a mass of strong ties that prevented him from speaking his mind freely on the union issue. Finally, Krackhardt showed how the complex network picture could be reduced to a simple but powerfully explanatory picture of the structure by performing a structural equivalence role analysis on Simmelian (co-clique) ties in the firm.

Network and Information Visualization: Patterns in Economic Networks

Lothar Krempel (Max Planck Institute for the Study of Societies, Cologne)

We interpret world trade data as a valued graph and use a spring embedder for the layout of the graph and explore how to map additional external information on to this graph. This allows to detect local concentrations and correlations of external information with the positions in the layout, and can greatly enhance the understanding of processes and properties of complex networks. It is the core of an explorative multivariate technology.
We show how to use color schemes for nominal and quantitative node attributes (geographic regions, GNP per capita) and how to derive meaningful colorschemes from these node attributes for the lines of the graph.

In a second part of the presentation we give examples for the study of empirical line attributes (growth rates for the flows, the composition of trade), which helps to identify regions of growth and decline, when such information is simultaneously aggregated for the nodes of the graph and communicated with zone symbols.

A Framework for Testing Social Network Visualization

Cathleen McGrath (Loyola Marymount University)

We have made great advances in techniques for conveying social network data using spatial positioning, node and edge characteristics, dimension, motion and user interaction. At its best, social network visualization has been directed by both principles of statistical graphical display and graph drawing aesthetics. In this work we undertake empirical analysis of viewer's perceptions of graphs based on differing layouts and the use of motion. We note that information can be extracted at at least three levels: graph theoretic (e.g. shortest path between two nodes); social network (betweenness centrality); and finally the social systems level (e.g. effects agents for change). We found that the best representation technique depends on the exact information that is to be conveyed.

Use of Node-Link Graph Analysis in Generating Instructions for Multiple Users

Margaret Mitchell (Macquarie University)

Directed acyclic graphs are a common method of representing groups of tasks whose underlying order is partial, having the advantage of easily representing nonlinear relationships, but not always an ideal one, as relationships between groups of tasks may go unnoticed as a group of tasks becomes large. Textual instructions have the advantage of allowing explicit reference to tasks or groups of tasks, but they do not facilitate agents working in parallel. We present the use of directed acyclic graph analysis as a means of overcoming these disadvantages. We use this as a means of text planning, producing instruction output that embodies the advantages of both graphical and textual representations.

Graphs with a Hierarchy – Models and Operations

Marcus Raitner (University of Passau)

There are several approaches to graphs with a hierarchy which superimpose a hierarchical structure on a graph. However there is no practical data-structure for these type of graphs. Here we try to give an overview of possible definitions and models of hierarchically structured graphs. This shall be taken as a starting point for an implementation of the data-structure.

Quality Evaluation of Search Engines

Marc Rittberger (University of Konstanz)

We present an information quality framework for measuring information services, especially search engines. We use the overall quality level information, communication & interaction, presentation, social & organization, and technique & method. Subcriteria for search engines of the information level are e.g. retrieval performance and meta information. The German search engines AltaVista.de, Lycos.de and Qualigo.de were used to evaluate retrieval performance. There is high evidence that Lycos.de performed significantly better on top 20 precision than the other search engines, because of the use of its catalogue data. For the presentation of meta-information we asked experts and novice user, which context information from German search engines result lists they liked most. The evaluation points pout that the number of answers, the title, the URL, and the highlighting are the most important ones. The experts and novice users judged significantly different on the importance for abstract, title, topic, URL, and query refinement.

Background on XML

Roberto Tamassia (Brown University)

A brief introduction to XML and related technologies is provided. This talk gives

background for the next presentation (by Ulrik Brandes) on XML standards for the graph drawing research community.

A Framework for the Visualization of Networks

Ioannis G. Tollis (University of Texas)

Visualizations of networks which show the inherent strengths and weaknesses of structures with clustered views are advantageous additions to many telecommunication network design tools. In this talk we present a framework for visualizing networks in a circular fashion. First we present an algorithm which lays out a biconnected network (graph) onto a single embedding circle. Furthermore, we can guarantee that if a zero crossing circular embedding exists for an input graph, then our algorithm will find it in linear time. Also, extensive experiments conducted over a set of 10,543 biconnected graphs show our technique to perform significantly better than previous techniques. Next we present algorithms for producing circular drawings of networks represented by nonbiconnected graphs. The drawings produced by these techniques clearly show the biconnectivity structure of the given networks. We also include results of an extensive experimental study which show our approach to significantly outperform the current state of the art.

Social Network Analysis and Algorithmic Visualization

Dorothea Wagner (Universität Konstanz)

We study algorithmic aspects of visualizing social networks in a way that supports the analysis. Our research is based on a systematic approach for network visualization following the general principles of information visualization. There are three aspects to distinguish, the network substance contained in the data, the design of the visualization defining the way substance is communicated and, finally, the algorithm realizing the design.
In the talk, we report on two examples where algorithms are developed to construct appropriate layouts of social networks. The first example consists in centrality layouts, i.e. layouts communicating the exact centrality values of actors while satisfying general criteria for readability. The algorithm proposed for cen-

trality layouts is evaluated using data from a study of local drug policy in Germany. In the second example, we study status layouts. We propose an algorithm that constructs layered layouts where the y-coordinates of the layers are chosen according to the exact status values in the network.
The talk is based on joint work with Ulrik Brandes.