Report from Dagstuhl Seminar 15052

Empirical Evaluation for Graph Drawing

Edited by

Ulrik Brandes[1], Irene Finocchi[2], Martin Nöllenburg[3], and Aaron Quigley[4]

1. **University of Konstanz**, ulrik.brandes@uni-konstanz.de
2. **University of Rome "La Sapienza"**, finocchi@di.uniroma1.it
3. **KIT – Karlsruhe Institute of Technology**, noellenburg@kit.edu
4. **University of St. Andrews**, aquigley@st-andrews.ac.uk

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 15052 "Empirical Evaluation for Graph Drawing" which took place January 25–30, 2015. The goal of the seminar was to advance the state of the art in experimental evaluation within the wider field of graph drawing, both with respect to user studies and algorithmic experimentation.

Seminar January 25–30, 2015 – http://www.dagstuhl.de/15052

1998 ACM Subject Classification E.1 Data Structures: Graphs and Networks, F.2 Analysis of Algorithms and Problem Complexity, G.3 Probability and Statistics: Experimental Design, H.5 Information Interfaces and Presentation

Keywords and phrases graph drawing, experimental design, algorithm engineering, user studies, empirical evaluation, information visualization

Digital Object Identifier 10.4230/DagRep.5.1.243

1 Executive Summary

Graph Drawing provides, among other things, the algorithmic foundations for network information visualization. It has considered implementation and experimentation as integral aspects from its very inception and recent research has demonstrated varying approaches to empirical evaluation. Experimental standards, however, have never been established, and little progress toward higher levels of sophistication can be observed.

The seminar was a community effort organized as a hands-on training event. It brought together experts on experimentation from fields with an established experimental tradition (referred to as "trainers"), and a group of graph drawing researchers expected to act as exponents and multipliers ("participants"). After two days of invited lectures on experimental methodology in different disciplines and a problem selection session, participants spent three days in working groups designing experiments. Trainers moving between groups and intermittent reporting session facilitated knowledge dissemination.

Participant feedback in the Dagstuhl survey indicates that the inclusion of trainers was highly appreciated. A number of experimental designs for a broad range of problems have been developed, and it is expected that many of them will be implemented and carried out in collaborative follow-up work.

As everyone who has ever been to Schloss Dagstuhl knows, Dagstuhl seminars are the ideal forum for achieving such goals. The fact that a considerable part of the graph drawing community came together for a week to focus on experimentation is expected to lead to a rapid diffusion of the seminar results and foster the acceptance of new methodology and criteria within the community.

On behalf of all participants, the organizer express their sincere gratitude to the Dagstuhl staff for their outstanding service and support.

2 Table of Contents

3 Organization

3.1 Introduction

Graph Drawing has a long tradition of implementing algorithms and evaluating their performance, maybe longer than other areas in algorithmics. An early example is the almost 20-year-old work of Himsolt [1], who evaluated twelve representative graph drawing algorithms based on a statistical analysis of geometric and structural properties of the respective layouts. Or the work of Di Battista et al. [2], who experimentally compared four algorithms for orthogonal grid drawings based on nine different quality measures. They introduced a benchmark suite, which is known today as the *Rome Graphs* and still frequently used in experimental graph drawing. The collection consists of more than 11,500 sparse graphs with fewer than 100 vertices generated from real-world graphs in software engineering and database systems.

This tradition may not come as a surprise given that Graph Drawing is a particularly interesting area for experimentation – an area that combines combinatorics, geometry, topology, algorithmics, visualization, interaction, and human factors. In this seminar, we are interested in two types of experiments which exhibit characteristics that are particularly challenging in graph drawing:

1. Experiments that compare graph-drawing algorithms in terms of domain-specific aesthetic optimization criteria (such as number of crossings, number of bends, angular resolution, crossings angles, layout area, uniformity of edge lengths, vertex distribution, or number of symmetries), and also in terms of running time and other more usual performance criteria.
2. Experiments that test how certain drawing styles help or hinder users to fulfill certain graph reading tasks. The difficulty of controlling for layout features sets the problem apart from other, more routinely conducted user studies in information visualization and human-computer interaction.

In spite of early and extensive work in these types of experiments, we think that it is time for the community to reconsider whether its experimental standards are up to date. We observe little progress in sophistication of experiments. One may ask whether this is because saturation has been reached early on but we doubt this. We simply think that knowledge about experimental methodology is not yet commonplace. Specifically, we identify the following problems:

1. In algorithmic studies, researchers often define the experimental region ad-hoc. They rarely ensure that the benchmark data is representative. Rather than generating good test instances or new benchmark sets, most researchers resort to the above-mentioned Rome graphs or to AT&T graphs, another benchmark suite.
2. It is rare that hypotheses are first explicitly formulated and then tested.
3. Phenomena that are observed during studies are usually explained post-hoc. Instead, such phenomena should lead to new hypotheses and to further experiments for validation.
4. There are only few specific graph generators.
5. Generally, there is a lack of user studies. Moreover, they tend to suffer from badly controlled factors in the instances presented to subjects.
6. Often, confounding factors and systematic biases are not identified.
7. Experiments are not convincingly randomized.
8. Statistical evaluation is generally rudimentary.

There were two main goals of this seminar. The first one was to increase the awareness of the need of high standards of empirical evaluations within the graph drawing community. We looked beyond graph drawing and learned from other fields with more advanced experimental research. Experimentation experts with experience in the closely related fields algorithm engineering, information visualization, and human-computer interaction, as well as experts from disciplines with more extensive experimental traditions who acted as trainers. By the close interaction with these experts, facilitated by a set of invited lectures and group discussions, we made a concerted effort for advancing the state of the art of experimental research in graph drawing. We aimed at establishing principles and experimental methodology by means of a guided knowledge import and an appropriate adaptation to the graph drawing context. Depending on progress with the second goal below, this shall result in a position paper.

The second goal was to actually design a set of empirical studies for answering experimental research questions in graph drawing. Within groups consisting of graph-drawing researchers as domain experts, and with experimentation experts floating among groups to offer advice, the newly acquired knowledge was applied to concrete problems. These problems had been collected in a special session following the invited tutorials. We hope that a large share of the ideas generated during this seminar will soon be implemented. Several groups and subgroups pledged to run some of the proposed experiments, evaluate them, and publish the results.

3.2 Schedule

As is customary, the schedule was organized around meals.

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
		breakfast		
welcome and introduction	invited talk discussion	topic selection group formation	trainer feedback working groups	working groups
		coffee		
invited talk discussion	topic collection	working groups	working groups	group reports final discussion
		lunch		
invited talk discussion		working groups	working groups	
		coffee & cake		
invited talk discussion	*excursion*	rapid feedback working groups	working groups	
		dinner		

3.3 Evaluation

According to the Dagstuhl survey conducted toward the end of the seminar, as well as informal feedback received by the organizers, the seminar was highly appreciated. Participants ranked the seminar higher than the (already hugely successful) average Dagstuhl seminar in virtually all dimensions, including scientific quality, group composition, the inspiration of new ideas for research and collaboration, and the acquisition of new knowledge. Increasing flexibility in the schedule and leaving more time for socializing are things to consider for future seminars.

Given the goals of the seminar it was by design that the group of participants had an overrepresentation of researchers who identify themselves as neither junior nor senior, and also happened to be relatively experienced Dagstuhl-goers, and an underrepresentation of industry.

The single most notable aspect in the qualitative feedback items was the excellent contribution of the trainers to the seminar.

We conclude that both the format and the content of the seminar worked as hoped for, despite adversarial circumstances that included the flu trying to compete with the spread of knowledge and new ideas.

References

1 Michael Himsolt. Comparing and evaluating layout algorithms within GraphEd. *J. Vis. Lang. Comput.*, 6(3):255–273, 1995.

2 Giuseppe Di Battista, Ashim Garg, Giuseppe Liotta, Roberto Tamassia, Emanuele Tassinari, and Francesco Vargiu. An experimental comparison of four graph drawing algorithms. *Computational Geometry*, 7(5–6):303–325, 1997.

4 Invited Presentations

We were fortunate to be joined by four enthusiastic and supportive domain experts in experimentation who fully embraced the goals of the seminar and immersed themselves into the group. One more trainer with a background in physics had to cancel on short notice.

4.1 The Art and Science of Evaluating Graph Layout Systems

Bernice E. Rogowitz (Visual Perspectives Research and Consulting – New York, US)

In graph layout evaluation, we measure the effectiveness of our computational methods and the degree to which our systems enable human effectiveness. In both cases, there are two evaluation strategies.

Methods	A-B Tests	Hypothesis-Generated Evaluations
Computational Evaluation		✓
Perceptual/Cognitive Evaluation		✓

One strategy is to evaluate performance using simple A-B tests, where design choices are compared for a particular set of stimuli using a particular task. In this paper, I make the case for hypothesis-driven evaluation, whose aim is to understand not only whether there is a difference between conditions, but why. In this more scientific approach, we select test conditions and tasks in order to test hypotheses about the underlying mechanisms driving observable distinctions, providing more generalizable results.

In this talk, I focus on perceptual and cognitive evaluation methods. Visualization systems, including graph layout methods, are designed to support human problem solving, judgment, decision-making, and pattern recognition. This means that every algorithm or tool we create embodies hypotheses about human information processing. Which perceptual and cognitive mechanisms are involved, however, depends on the task. For example, detection and discrimination tasks involve early vision and tasks that require finding features in data require attention, semantic encoding, and memory. It is important to match the task to the perceptual and cognitive demands of the application.

I show examples of how evaluation experiments can be used to suggest new parameters for algorithms, test the ecological validity of different data representation schemes, and measure differences between intuitive and objective measures of visualization effectiveness. For example, I describe an experiment with Frank van Ham that shows how Gestalt principles of perceptual organization can enhance graph layout algorithms. When observers were allowed to move the nodes of a planar graph, they created configurations that emphasized clusters in the data. Their results looked quite similar to a force-minimization layout, but with less-uniform edge lengths, fewer line crossings, and increased distance between clusters. They even used edges to create convex hulls around clusters, enhancing their structure.

In an experiment with Mercan Topkara, Arum Hampapur and Bill Pfeiffer, I show how visual evaluation methods can be used to select the hit and false alarm rate of a surveillance algorithm to maximize human performance. And, in an experiment with Enrico Bertini, Aritra Dasgupta, Jorge Poco and Claudia Silva, I demonstrate how judgments of data magnitude and spatial distribution can be influenced by the choice of color scales. Surprisingly, the color scale with highest appeal, judged accuracy and familiarity provided the worst performance, demonstrating the importance of empirical testing.

4.2 Controlled Experiments in Software Engineering

Janet Siegmund (Universität Passau, DE)

Empirical research in psychology has come a long way. Thus, computer scientists who want to evaluate the human factor in their discipline, e.g., graph drawing or software engineering, can profit from the methodological advances in psychology. In this talk, I present a roadmap for conduction empirical studies of the human factor related to computer science research objectives. This roadmap is based on the state of the art of empirical research in psychology.

4.3 Designing Experiments in Political Science

Michael Stoffel (Universität Konstanz, DE)

This talk introduced the so-called "potential outcomes" framework that is the foundation of experimental research. Building on this, we discussed the two general assumptions that experimenters make and how to guarantee that they are satisfied: unconfoundedness and the stable unit treatment value assumption (SUTVA). In the empirical part of the talk, we then had a look at an experiment on principal-agent relations in the bureaucracy.

4.4 Experimental Algorithmics

Catherine C. McGeoch (Amherst College and D-Wave Systems Inc., US)

I talk about experimental methods that are aimed at algorithmic questions. These are more likely to be descriptive and exploratory (especially graphical) than confirmatory in nature – due to the common types of questions asked in algorithm studies. I review some possibilities for choosing performance indicators and show how variance reduction techniques can improve outcomes.

5 Working Groups

Initial ideas for topics to be worked on in groups were collected in the discussions following each invited presentation. The topics were reviewed, complemented, and consolidated in a special session at the end of the tutorial part. Self-assignment of participants to groups was surprisingly easy to deliberate.

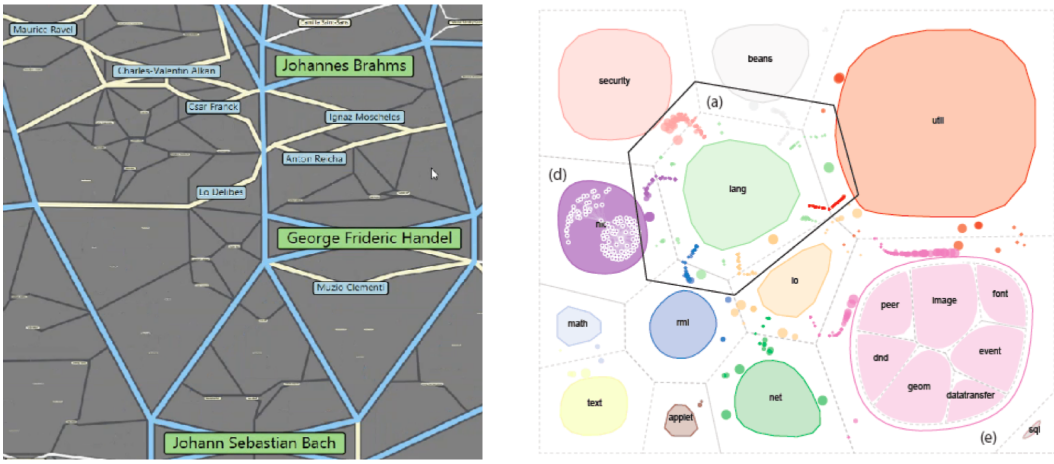
Working groups used the Dagstuhl Wiki environment to collect input, ideas, and outcomes, and the following are reports distilled from these entries.

5.1 Large Graphs

Irene Finocchi, Seokhee Hong, Lev Nachmanson, Huamin Qu, Alexander Wolff, and Kai Xu

Due to the finite resolution of display devices, which represents a physical limitation on the size of graphs that can be conveniently displayed, designing effective visual representations of large graphs poses many challenges. In order to cope with the cluttering effects arising when visualizing huge quantities of data, it seems important to use information hiding techniques and decompositions of the visual space that reflect some structural view of the data. The working group focused on two different visualization paradigms, both inspired by graph maps, that appear to be promising when dealing with large graphs. The implementation of these paradigms – called *graph maps with highways* and *clustering with maps*, respectively – was available in software tools co-designed by two of the group participants. For each paradigm, a different user study has been designed identifying both hypotheses and tasks to be performed by end users. The goal of the user studies is to compare the effectiveness of drawings produced by different algorithms according to the specific tasks, highlighting for which tasks each paradigm turns out to be most useful.

Graph maps with highways. For many real-world graphs with substantial numbers of edges, traditional algorithms produce visually cluttered layouts [3]. The relations between the nodes are difficult to analyze by looking at such layouts. Graph maps with highways, designed by Nachmanson *et al.*, exploit techniques similar to edge bundling to solve this problem. A visualization example of a graph map with highways is shown in Figure 1(a). The hypothesis of the user study is that the highway metaphor improves node location *memorability*, i.e.,



(a) **(b)**

Figure 1 (a) Graph map with highways on the composers' network from [1]; (b) Clustering with maps: an example taken from [2].

the speed of users to find previously visited nodes. During the study, each user must visit a given number of nodes, using pan & zoom, and then find one of the previously visited nodes as quickly as possible. Data used for the experiment include social networks and the composers' network used in the Graph Drawing 2011 competition [1].

Clustering with maps. This visualization paradigm, described in [2], has been designed with the goal of representing the major communities in large social networks. A visualization example is shown in Figure 1(b). In the designed study, end users are required to perform either inter-cluster tasks (e.g., counting clusters or finding the "most related" pair of nodes) or intra-cluster tasks based on zooming (e.g., finding an opinion leader inside a certain cluster). The study should test the following three hypotheses:

- a reasonable gap between the clusters increases the speed of counting clusters;
- if there are many nodes belonging to two different clusters, the gap between those clusters in the visualization should be larger;
- a big gap is good for inter-cluster tasks, while a small gap is more convenient for intra-cluster tasks.

References

1. Christian A. Duncan, Carsten Gutwenger, Lev Nachmanson, and Georg Sander. Graph drawing contest report. In *Graph Drawing – 19th International Symposium, GD 2011*, pages 449–455, 2011.
2. Yanhong Wu, Wenbin Wu, Sixiao Yang, Youliang Yan, and Huamin Qu. Interactive visual summary of major communities in a large network. In *IEEE Pacific Visualization Symposium, PacificVis 2015*, 2015. To appear.
3. Sergey Pupyrev, Lev Nachmanson, Sergey Bereg, and Alexander E. Holroyd. Edge routing with ordered bundles. In *Graph Drawing*, volume 7034 of *Lecture Notes in Computer Science*, pages 136–147. Springer, 2012.

5.2 Bends, Curves, and Bundles

Daniel Archambault, Martin Fink, Martin Nöllenburg, Yoshio Okamoto, and Ignaz Rutter

Edge bundling is a popular technique for reducing visual clutter in layouts of dense graphs [1, 2, 3, 4, 5, 6]. It is based on the idea of grouping edges whose end-vertices have similar locations into *bundles* and using appropriate graphical deformations to draw the edges of each bundle along the same underlying trunk path. While bundling techniques are claimed to successfully reduce edge clutter, the effect of bundled layouts on human graph readability is not yet well investigated and only very few studies have been published [7]. It may be argued that bundled layouts represent global trends well on a coarser scale, but there is also an unavoidable trade-off between cleaning up the layout by edge bundling and losing low-level connectivity information in the graph. This is because in an edge bundle it is often very difficult to trace individual edges and a dense bundle may be indistinguishable from a complete bipartite subgraph linking all pairs of vertices on both sides of the bundle. This is the main difference to the graph drawing style of *confluent layouts* [8]. In a confluent layout of a graph $G = (V, E)$, edges merge and split in smooth confluent junctions such that there is an edge between two vertices in G if and only if there is a smooth path between them in the layout.

In this working group we set out to design an empirical user study to evaluate the influence of edge bundling strength on typical graph reading tasks, both of global nature and detail-oriented ones. We hypothesize that edge bundling has a positive effect on tasks that require more global reasoning about the layout, but that it has a negative effect on tasks that require detailed knowledge about local structures of the graph. In order to measure and control bundling strength on a continuous scale, we defined an ambiguity measure as well as a measure for the amount of edge deformation. We plan to evaluate the task performance for four basic tasks that have been extracted from practical applications of graph visualization. Graph data will contain both geographic networks with fixed vertex positions as well as force-directed layouts of non-spatial networks, e.g., social networks. Since our aim is to evaluate edge bundling as a general technique and not a particular bundling algorithm, we will focus on a few representative algorithms that create explicit edge bundles so that we can vary the bundling strength by interpolating between the unbundled input layout and the fully bundled layout as computed by the selected algorithms. This necessarily excludes bundling methods that include explicit edge disambiguation techniques [9, 10].

Currently, in preparation of the planned user study, we are implementing an interpolation feature for varying the bundling strength into our selected edge bundling algorithms and we are collecting suitable real-world data sets for the study.

References

1 Weiwei Cui, Hong Zhou, Huamin Qu, Pak Chung Wong, and Xiaoming Li. Geometry-based edge clustering for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1277–1284, Nov 2008.
2 Danny Holten and Jarke J. van Wijk. Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, 28(3):983–990, 2009.
3 A. Lambert, R. Bourqui, and D. Auber. Winding roads: Routing edges into bundles. *Computer Graphics Forum*, 29(3):853–862, 2010.

4 E.R. Gansner, Yifan Hu, S. North, and C. Scheidegger. Multilevel agglomerative edge bundling for visualizing large graphs. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE*, pages 187–194, March 2011.
5 Quan Nguyen, Seok-Hee Hong, and Peter Eades. TGI-EB: a new framework for edge bundling integrating topology, geometry and importance. In Marc van Kreveld and Bettina Speckmann, editors, *Graph Drawing*, volume 7034 of *Lecture Notes in Computer Science*, pages 123–135. Springer Berlin Heidelberg, 2012.
6 Hong Zhou, Panpan Xu, Xiaoru Yuan, and Huamin Qu. Edge bundling in information visualization. *Tsinghua Science and Technology*, 18(2):145–156, April 2013.
7 Fintan McGee and John Dingliana. An empirical study on the impact of edge bundling on user comprehension of graphs. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI'12, pages 620–627, New York, NY, USA, 2012. ACM.
8 Matthew Dickerson, David Eppstein, Michael T. Goodrich, and Jeremy Y. Meng. Confluent drawings: Visualizing non-planar diagrams in a planar way. *Journal of Graph Algorithms and Applications*, 9(1):31–52, 2005.
9 Sheng-Jie Luo, Chun-Liang Liu, Bing-Yu Chen, and Kwan-Liu Ma. Ambiguity-free edge-bundling for interactive graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):810–821, May 2012.
10 Quirijn Bouts and Bettina Speckmann. Clustered edge routing. In *Pacific Visualization Symposium (PacificVis'15)*. IEEE, 2015. To appear.

5.3 Cognition

Rudolf Fleischer, Stephen G. Kobourov, Tamara Mchedlidze, Wouter Meulemans, Aaron Quigley, and Bernice E. Rogowitz

The cognition group started its graph drawing [1, 2] work asking a series of broad questions before digging down into specific details. We started with questions relating to aesthetic experience [3, 4, 5] such as "What is the purpose of the visualisation?," "analysis vs. communication," "exploratory vs confirmatory vs communication," "is it to tell a message?" or "is this to allow people to explore?" we moved onto discussing how visualisations can be artist [6], engaging, beautiful [7], attractive [8] and even perhaps arresting to draw people into use not just for an immediate short term reaction but instead a long term use. The question of making a visualisation arresting raises a number of questions including, does familiarity with a visual language breed contempt, is this a suitable goal, and does the layout or rendering impact the arresting nature of a display overall. If a visualisation can be arresting how does this affect long term use and memorability for ongoing use. Many of the visual affects discussed relate to the notion of a "honeypot effect" which maybe enough to draw someone into use but not long term engagement.

Next, the working group moved onto the discussion of static versus dynamic displays and interactive versus non-interactive displays. The notion of making a visualisation arresting is interlinked with is the displayed content static or dynamic. Further, does an arresting visualisation draw someone in with the expectations they might transition into interactive engagement? The design space moving from initial engagement into these factors opens up a large space.

The combination of this broad design space with the goal of hooking users in with arresting visualisations require careful thought on the measures [9] we can employ in the evaluation of suitability. The group moved onto a specific goal of exploring clusters in force directed graph drawings and what are the key aspects of visual appeal and performance. This brings into questions of shape, colour harmony, symmetry, geometry [10, 11] and gestalt principles [12, 13, 14, 15].

The outcome from the group is the design for an experiment which can be run in a lab and also online. A set of hypothesis around, appeal, visual principles, familiarity, subjective measures, display properties, geometric principles [16] have been formed. From this a series of experiments with small world graphs will measure, attraction, engagement and memorability [17, 18].

Our methods are based on the generation of drawings of small world-type graphs, a set of independent measures, dependent measures with various conditions and over 200 stimuli.

References

1 Frank van Ham and Bernice Rogowitz. Perceptual organization in user-generated graph layouts. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1333–1339, Nov 2008.

2 Miro Spönemann, Björn Duderstadt, and Reinhard von Hanxleden. Evolutionary meta layout of graphs. In Tim Dwyer, Helen Purchase, and Aidan Delaney, editors, *Diagrammatic Representation and Inference*, volume 8578 of *Lecture Notes in Computer Science*, pages 16–30. Springer Berlin Heidelberg, 2014.

3 Chris Bennett, Jody Ryall, Leo Spalteholz, and Amy Gooch. The aesthetics of graph visualization. In *Proceedings of the Third Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'07, pages 57–64, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.

4 A. Lau and A. Vande Moere. Towards a model of information aesthetics in information visualization. In *Information Visualization, 2007. IV'07. 11th International Conference*, pages 87–92, July 2007.

5 Edward Vessel, Gabrielle Starr, and Nava Rubin. The brain on art: intense aesthetic experience activates the default mode network. *Frontiers in Human Neuroscience*, 66(6), 2012.

6 Werner Vogels. Graphs as art, 2006. Blog post, http://www.allthingsdistributed.com/2006/07/graphs_as_art.html.

7 Edward Vessel and Nava Rubin. Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, 10(2), 2010.

8 Julie Steele and Noah Iliinsk. *Beautiful Visualization: Looking at Data Through the Eyes of Experts.* 1st ed. O'Reilly Media, 2010.

9 Colin Ware, Helen Purchase, Linda Colpoys, and Matthew McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2):103–110, June 2002.

10 Moshe Bar and Maital Neta. Humans Prefer Curved Visual Objects. *Psychological Science*, 17:645–648, 2006.

11 Roman Chernobelskiy, Kathryn I. Cunningham, Michael T. Goodrich, Stephen G. Kobourov, and Lowell Trott. Force-directed lombardi-style graph drawing. In Marc van Kreveld and Bettina Speckmann, editors, *Graph Drawing*, volume 7034 of *Lecture Notes in Computer Science*, pages 320–331. Springer Berlin Heidelberg, 2012.

12 Nadia Ali and David Peebles. The effect of gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. *The Journal of the Human Factors and Ergonomics Society*, 55(1):183–203, Feb 2013.

13 Marlies de Brouwer. The influence of the gestalt principles similarity and proximity on the processing of information in graphs: An eye tracking study. Master's thesis, Tilburg University, 2014. http://arno.uvt.nl/show.cgi?fid=134020.
14 C.T. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *Computers, IEEE Transactions on*, C-20(1):68–86, Jan 1971.
15 A. Rusu, A.J. Fabian, R. Jianu, and A. Rusu. Using the gestalt principle of closure to alleviate the edge crossing problem in graph drawings. In *Information Visualisation (IV), 2011 15th International Conference on*, pages 488–493, July 2011.
16 A. van Goethem, W. Meulemans, B. Speckmann, and J. Wood. Exploring curved schematization. In *2014 IEEE Pacific Visualization Symp. (PacificVis)*, pp. 1–8, March 2014.
17 Scott Bateman, Regan Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *ACM Conference on Human Factors in Computing Systems (CHI 2010)*, pages 2573–2582, Atlanta, GA, USA, 2010. Best paper award.
18 Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVis 2013)*, 2013.

5.4 Computational Experiments

Ulrik Brandes, Emilio Di Giacomo, Andreas Karrenbauer, Karsten Klein, and Maurizio Patrignani

The group decided to design computational experiments for one specific research question. The question was identified based on various criteria such as relevance to the field of graph drawing, difficulty to be addressed analytically, expertise of the group members, and likelihood of leading to an actual study. This is what we came up with:

Why do force-directed layouts exhibit relatively few crossings?

The approach taken to address this question is to fix a layout algorithm, identify structural features that are drawn with a crossings in optimal layouts, and relate the occurrence of such features in the input to crossings in the output.

Theoretical considerations allowed us to identify several families of problematic subgraphs, even though we conjecture that there exist trees without any of these subgraphs that still result in crossings. Based on these insights we derived four concrete hypothesis that are sufficiently specific to be testable.

Fortunately, these hypotheses give us a reason to test on planar graphs as for them all observed crossing are caused by the layout algorithm. A particularly important observation was made only because of the preceding discussion on experimental design: instead of simply charting the number of crossings we will determine the matches of problematic subgraphs and crossings, because these provide the evidence whether the crossings are actually caused by assumed mechanism.

Current challenges include fixing the set of problematic subgraphs to study, proving the above conjectures, random sampling of planar graphs, efficient counting of subgraphs, and identifying and controlling random and systematic biases due to the layout algorithm implementation.

5.5 Experiments Involving Humans

Markus Chimani, Walter Didimo, Michael Kaufmann, Giuseppe Liotta, and Dorothea Wagner

The group's main goal is to design an experiment centered around human understanding of graph drawings. A main goal thereby is to devise an experimental setup based on the fundamentals discussed in this workshop, to avoid traditional shortcomings often found in GD user studies. The study should hence not be an afterthought to theoretical research, but spur interest in and shed light on a topic that is typically rather left to intuition than to scientific rigor. Our naïve sounding question is:

Where should we put the arrow heads in directed graph drawings?

Despite the fact that we all draw directed graphs on a day-to-day basis, and that most (but not all) of us tend to draw the arrow heads at the edges' ends, it is unclear if this drawing paradigm is in fact the most suitable one. There are previous user studies discussing the drawing of directed edges, devising very different and diverse drawing approaches (even animated ones, unsuitable for printouts).

In contrast to those, we want to stick to the traditional arrow head paradigm, as there seems to be a large consensus that this paradigm is the most natural. However, when thinking about a vertex with large degree, traditionally end-placed arrow heads will overlap, making it hard or even impossible to identify the direction of a specific edge. We consider multiple different placement strategies to mitigate these effects. Some of these strategies give rise to interesting combinatorial optimization problems. However, our study will not discuss this; its outcome may, however, help to understand whether a detailed theoretical investigation of such a placement paradigm is at all worthwhile.

We spent a large percentage of our time at Dagstuhl devising our hypotheses and the tasks given to the users, discussing their interplay, and trying to find an as small set of tasks as possible, while still covering all our hypotheses in a meaningful way. Especially this minimization – necessary to end up with a feasibly conductible experiment – turned out to be harder than anticipated.

A further important fact for our user study is to specify our underlying graphs and drawings, which are to be based on real-world scenarios but yet controllable enough for a user study. Detailed discussions have taken place with respect to the various confounding factors, the experimental design and setup, and a time-plan for the next steps, including pilot studies to hammer out the finer choice details prior to the main experiment. The final steps towards the experimental study are currently in progress.

5.6 Tasks Linked to Representations

Tim Dwyer, Tamara Munzner, Falk Schreiber, Bettina Speckmann, and Matt F. Stallmann

In discussing "Experimental Graph Drawing," we felt that before a meaningful experiment could be designed, it was necessary to understand what the most important tasks in graph drawing and network visualization really are. Once the important tasks and challenges facing network analysts are properly understood then we can begin to test various methods for creating visualizations of networks that actually support those tasks.

We began with a cursory literature search for task surveys and taxonomies, we found a couple but not that were very deep. Lee et al. [1] give a brief and rather overview of some reasonable sounding tasks, particularly *low-level tasks* such as path following, common neighbours, etc. However, high-level tasks and in-particular, how these translate into real-world problem solving we felt was missing. Pretorius et al. [2] give a longer discussion of tasks for *multivariate network analysis*, however, again it seems unclear precisely when these tasks translate into "aha moments" in analysis. Munzner [3] gives a "how/what/why" framework for problem solving with graph visualization that perhaps gets closer to consideration of real-world applications, but this part of the book is much briefer in relation to network visualization than other types of data visualization.

In summary then, we feel that there is room for a deeper analysis of tasks starting from applications and working down to a generalizable taxonomy. Such a taxonomy should identify not only very low-level tasks (ones that could be considered atomic, e.g. "is node A connected to node B?" but also mid-level tasks that are very application agnostic but composed of multiple low-level operations (e.g., "what is the shortest path between A and B?") and high-level tasks more specific to applications (e.g., "what is the critical path in this workflow, which deadlines can slip without jeopardizing the entire project?" Armed with such a taxonomy we are ready to consider how visualization can help or even if it is always the best method for particular scenarios.

The example above is quite a concrete connectivity task that requires close inspection of precise connectivity information, but it also seems clear that people are interested in understanding much larger-scale graph structure. Examples of applications where the practitioners want some insight into the gross structure of very large networks with thousands, tens-of-thousands or even millions of nodes abound. For example, in Biology metabolic networks contain thousands of nodes and gene activity correlation networks contain tens-of-thousands of nodes, in neuroscience neural networks, economics... really any complex system considered by modern scientists and other analysts has a scale issue. Much work has been done in making algorithms and rendering processes scale to thousands or millions of nodes and links in an efficient way. Less well understood is exactly how these large-scale visualizations help practitioners, especially when large, naturally occurring networks when rendered as node link diagrams, tend to appear as "hairballs" or when rendered as matrices, as "white noise".

On this note, our discussion then diverged to consideration of alternatives for understanding gross network structure that might avoid many of these drawing pitfalls entirely. We discussed the possibility of computing summary statistics for graphs that can convey the high-level structure of networks concisely yet adequately to give practitioners the information they need about large networks. For example, a force-directed layout of a very large network may turn out to be a hairball which only tells you that the network has many nodes and links and it is likely small-world and scale free. However, you can't be sure of even these properties without further investigation of the node-degree distribution and the network diameter. So, why not by-pass the precise node-link diagram entirely and make the first visualization be a succinct dashboard display of statistics such as these?

This led to an extensive study of network-diagnostic statistics, or *NetNostics* or even *NetGnostics*. We found that there is extensive literature and the theory and practice of statistical analysis of networks, a book by Kolaczyk [4] gives a good overview and introduction.

Yet, such statistics are rarely the focus of visualization – particularly not in the "dashboard" view we envisage.

In summary then, we plan to proceed on two fronts. First, we will continue to work towards a deeper survey and taxonomy of tasks which we feel will be an important practical contribution to the emerging field of experimental graph drawing in giving a solid motivation and foundation for designing studies with ecological validity. Second, we will survey the field of statistical analysis of network structure – with a short term goal of publishing a survey that is useful to information visualization researchers but in the longer term, will would like to produce a practical "dash-board" system as described above.

William Hill and Bernice Rogowitz contributed to subsequent discussion.

References

1. Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV'06, pages 1–5, New York, NY, USA, 2006. ACM.
2. A. Johannes Pretorius, Helen C. Purchase, and John T. Stasko. Tasks for multivariate network analysis. In *Multivariate Network Visualization*, volume 8380 of *Lecture Notes in Computer Science*, pages 77–95. Springer-Verlag, 2014.
3. Tamara Munzner. *Visualization Analysis and Design.* A.K. Peters visualization series. CRC Press, 2014.
4. Eric D. Kolaczyk. *Statistical Analysis of Network Data.* Springer-Verlag, 2009.

Participants

- Daniel Archambault
Swansea University, GB
- Ulrik Brandes
Universität Konstanz, DE
- Markus Chimani
Universität Osnabrück, DE
- Emilio Di Giacomo
University of Perugia, IT
- Walter Didimo
University of Perugia, IT
- Tim Dwyer
Monash Univ. Melbourne, AU
- Martin Fink
University of California – Santa Barbara, US
- Irene Finocchi
University of Rome 'La Sapienza", IT
- Rudolf Fleischer
German University of Technology – Oman, OM
- Seok-Hee Hong
The University of Sydney, AU
- Andreas Karrenbauer
MPI für Informatik – Saarbrücken, DE
- Michael Kaufmann
Universität Tübingen, DE
- Karsten Klein
Monash University, AU
- Stephen G. Kobourov
Univ. of Arizona – Tucson, US
- Giuseppe Liotta
University of Perugia, IT
- Catherine C. McGeoch
D-Wave Systems Inc. & Amherst College, US
- Tamara Mchedlidze
KIT – Karlsruher Institut für Technologie, DE
- Wouter Meulemans
Universität Münster, DE
- Tamara Munzner
University of British Columbia – Vancouver, CA
- Lev Nachmanson
Microsoft Corp. – Redmond, US
- Martin Nöllenburg
KIT – Karlsruher Institut für Technologie, DE
- Yoshio Okamoto
The University of Electro-Communications – Tokyo, JP
- Maurizio Patrignani
Roma Tre University, IT
- Huamin Qu
HKUST – Kowloon, HK
- Aaron Quigley
University of St. Andrews, GB
- Bernice E. Rogowitz
Visual Perspective – New York, US
- Ignaz Rutter
KIT – Karlsruher Institut für Technologie, DE
- Falk Schreiber
Monash University, AU
- Janet Siegmund
Universität Passau, DE
- Bettina Speckmann
TU Eindhoven, NL
- Matthias F. Stallmann
North Carolina State Univ., US
- Michael Stoffel
Universität Konstanz, DE
- Dorothea Wagner
KIT – Karlsruher Institut für Technologie, DE
- Alexander Wolff
Universität Würzburg, DE
- Kai Xu
Middlesex University, GB