Report from Dagstuhl Seminar 21152

Multi-Level Graph Representation for Big Data Arising in Science Mapping

Edited by

Katy Börner[1] and Stephen Kobourov[2]

1 **Indiana University – Bloomington, US,** `katy@indiana.edu`

2 **University of Arizona – Tucson, US,** `kobourov@cs.arizona.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 21152 "Multi-Level Graph Representation for Big Data Arising in Science Mapping." The seminar brought together researchers coming from information visualization, psychology, cognitive science, human-computer interaction, graph drawing, computational geometry, and cartography with interests in the "science of science" to discuss novel graph mining and layout algorithms and their application to the development of science mapping standards and services. Due to the pandemic, this was a "hybrid" event with only 5 in-person participants and 25 by-zoom participants from over ten different countries and at least five different time-zones. There were three overview talks and four special presentations (evening webinars) from different communities represented in the seminar. Abstracts of these talks and presentations are collected in this report. Three working groups formed around open research problems related to the seminar topic and we report about their findings.

Seminar April 11–16, 2021 – http://www.dagstuhl.de/21152

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms; Human-centered computing → Visualization design and evaluation methods

Keywords and phrases graph theory, graph algorithms, science of science, cartography and map-like representations, network visualization, multi-level semantic zooming

Digital Object Identifier 10.4230/DagRep.11.3.1

1 Executive Summary

Katy Börner, Indiana University, USA, katy@indiana.edu
Stephen Kobourov, University of Arizona, USA, kobourov@cs.arizona.edu

Networks are all around us. At any moment in time we are driven by, and are an integral part of, many interconnected and dynamically changing networks. Our species has evolved in the context of diverse ecological, biological, social, and other types of networks over thousands of years. We created telephone and power networks, road and airline networks, the Internet and the World Wide Web. We model biological processes with metabolic and protein networks, fake news and rumors with epidemic networks, and even the brain with neural networks. The study of such large-scale networks is one of the prominent generators of big data. Analyzing, exploring, and understanding these complex, interdependent, multi-level networks requires new, more efficient, and more intuitive graph analysis and visualization approaches. This is confirmed in a 2017 VLDB study by Sahu et al. about how graphs are used in practice, where analysts identified scalability and visualization as the most important issues to address.

Multi-Level Graph Representation for Big Data Arising in Science Mapping, *Dagstuhl Reports*, Vol. 11, Issue 03, pp. 1–15
Editors: Katy Börner and Stephen G. Kobourov
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Recent advances in data, algorithms, and computing infrastructures make it possible to map humankind's collective scholarly knowledge and technology expertise by using topic maps on which "continents" represent major areas of science (e.g., mathematics, physics, or medicine) and zooming reveals successively more detailed subareas. Basemaps of science and technology (S&T) are generated by analyzing citations links between millions of publications and/or patents. "Data overlays" (e.g., showing all publications by one scholar, institution, or country or the career trajectory of a scholar as a pathway) are generated by science-locating relevant publication records based on topical similarity. Science maps are widely used to compare expertise profiles, to understand career trajectories, and to communicate emerging areas. The recent National Academy of Science Colloquium on *Modeling and Visualizing Science Developments* co-organized by Börner showcased the utility of predictive modeling and large-scale mapping efforts, e.g., in support of ranking institutions, analyzing job market developments, and innovation diffusion and technology adoption.

Despite the demonstrated utility of large-scale S&T maps, current approaches do not scale to the hundreds of millions of data records now available. Most users have a hard time reading large-scale networks and few can traverse or derive knowledge from multi-level presentations of networks. Most maps of science, technology, or jobs data support exactly one level of detail. Very few even support two levels such as the UCSD map of science. A key challenge is designing efficient and effective methods to visualize and interact with more than 100 million scholarly publications at multiple levels of resolution.

Given results from our prior studies on the effectiveness and memorability of map-like visualization of large graphs, we are interested to bring together leading experts to design a multi-level, large-scale map of science that can be used by experts and the general public alike.

The notion of multiple-levels-of-detail graph representation can be captured with Multi-Level Graph Sketches (MLGS) that take a static map-like representation to a multi-level setting needed for exploring and interacting with large, real-world networks. Using the familiar map metaphor, multi-level graph algorithms can make it possible to identify important nodes, major pathways, and clusters across multiple levels. Specifically, we aim to develop efficient algorithms with theoretical and practical guarantees for creating Multi-Level Graph Sketches (MLGS) in support of visual analytics tasks for large network exploration, navigation, and communication. Unlike existing methods for visualizing multi-level networks based on meta-nodes and meta-edges, the MLGS approach can provide real nodes (prototypes) and real paths (backbones) for each level, similar to geographic maps that show real cities and real roads at every level of detail.

Research questions included: (1) designing efficient algorithms for MLGS: advance the state-of-the-art in graph algorithms by generalizing the notion of graph spanners to multiple levels. (2) utilizing MLGS algorithms in visualization: applying the MLGS representation in the context of network analysis and visualization for interacting with large networks, which combines the MLGS approach with clustering, layout, and map-like visualization. (3) developing a new approach for science classification, lookup, and topical mapping service in support of data-driven decision making by students, teachers, and administrators. (4) validating the new algorithms and visualizations: evaluation based on quantitative metrics such as efficiency/scalability, stress/distortion and precision/recall, as well as qualitative metrics based on human subject studies of the utility of visualizations along with readability, engagement, and memorability.

The main goal of this seminar was to bring together researchers coming from information visualization, psychology, cognitive science, human-computer interaction, graph drawing, computational geometry, cartography, and GIS with interests in "science of science" to discuss

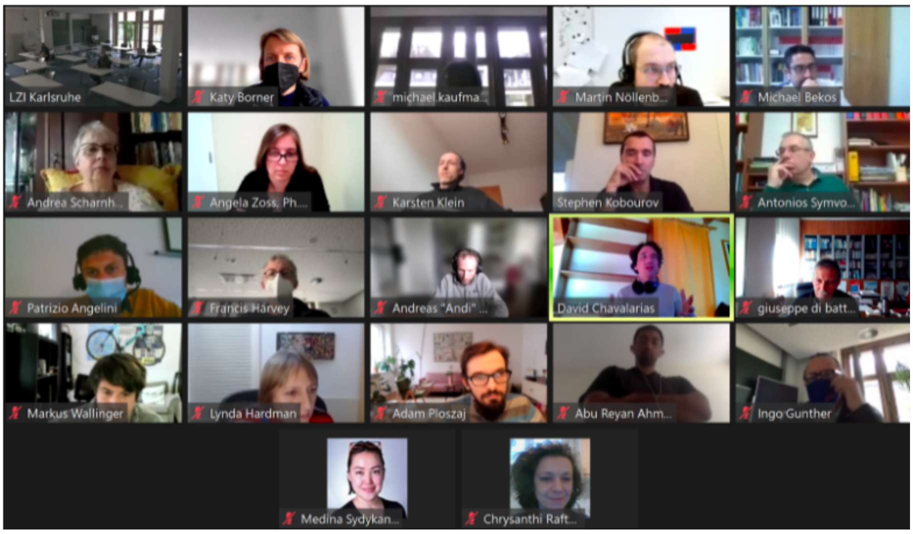

Figure 1 A very different look for the seminar...

novel graph mining and layout algorithms and their application to the development of science mapping standards and services.

Due to the pandemic, we had a "hybrid" seminar with only 5 in-person participants and 25 by-zoom participants; see Fig. 1. With participants from more than 10 different countries and at least 5 different time-zones (some as much as 9 hours behind Central Dagstuhl time), this was a new experience for most of us and definitely different from previous Dagstuhl seminars. Nevertheless, we attempted to adapt by having additional evening events and moving the traditional Wednesday excursion to the morning. As this was a highly interdisciplinary seminar, we started the event with talks that introduced the state of the art in the participating fields on a high level. After the introductory presentations, we presented our initial research problems and discussed further questions in an open-problem session.

A set of 4-6 research problems was then finalized and the formation of working groups for these problems was completed by the end of the second day. The remaining three days were dedicated to working-group meetings, progress reports, and initial write-ups.

The main expected outcome of this seminar will be a special issue of the journal *IEEE Computer Graphics and Applications*[1] on the main topics of the seminar. Specifically, we expect 4-6 research papers on the problems discussed by the working groups. Longer term goals include: improved collaborations and communications between the different communities brought together for this seminar, improved maps of science for SciMap2020, and new multi-level graph algorithms and approaches.

[1] `https://www.computer.org/digital-library/magazines/cg/call-for-papers-special-issue-on-multi-level-graph-representations-for-big-data-in-science`

2 Table of Contents

3 Overview of Talks and Special Events

3.1 The Value and Need for Mapping Science

Katy Börner (Indiana University – Bloomington, US)

In the information age, the ability to read and make data visualizations is as important as the ability to read and write. This talk explains and exemplifies the power of data visualizations and the need for multi-level network visualizations. It introduces a theoretical visualization framework meant to empower anyone to systematically render data into insights together with tools that support temporal, geospatial, topical, and network analyses and visualizations. Materials from the *Visual Analytics* course (`https://visanalytics.cns.iu.edu`) and science maps from the *Places & Spaces: Mapping Science* exhibit (`http://scimaps.org`) will be used to illustrate key concepts. In addition two projects are introduced: (1) SPOKE, a graph-theoretic database that supports biomedical researchers in the exploration of 30 different public datasets comprising 3 million edges and 30 million edges, enabling new discoveries and (2) HuBMAP, an NIH-funded project that aims to map the human body at the single-cell level.

References

1 1 Katy Börner. 2021. Atlas of Forecasts: Modeling and Mapping Desirable Futures. Cambridge, MA: The MIT Press.

2 Weber Griffin, Yingnan Ju, and Katy Börner. 2020. "Considerations for Using the Vasculature as a Coordinate System to Map All the Cells in the Human Body". Frontiers in Cardiovascular Medicine 7 (29): doi: 10.3389/fcvm.2020.00029.

3 Michael Snyder, et al. 2019. "The human body at cellular resolution: the NIH Human Biomolecular Atlas Program". Nature 574: 187-192.

4 Katy Börner, Andreas Bueckle, and Michael Ginda. 2019. "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments". PNAS, 116 (6) 1857-1864.

5 Katy Börner. 2015. Atlas of Knowledge: Anyone Can Map. Cambridge, MA: The MIT Press.

6 Katy Börner. 2010. Atlas of Science: Visualizing What We Know. Cambridge, MA: The MIT Press.

3.2 Multi-Level Graph Representations

Stephen Kobourov (University of Arizona – Tucson, US)

There are many algorithms, tools, and online services that support the analysis and visualization of networks. However, to the output of most large-scale, multi-level network visualizations often resembles either hairballs or counter-intuitive hierarchical cluster representations (via high-level meta-nodes and meta-edges) which use "super-noding" and "edge bundling." Here, we describe a multi-level graph representation that takes the static map-like representations to a multi-level setting needed for exploring and interacting with large, real-world networks. Using the familiar map metaphor, multi-level graph algorithms can make it possible to identify important nodes, major pathways, and clusters across multiple levels.

3.3 How can our empirically evaluated design guidelines and metaphors help you depict aesthetically pleasing, engaging, societally relevant, perpetually salient, and cognitively supportive multi-level graph networks of big data

Sara Irina Fabrikant (University of Zürich, CH)

The choice of appropriate colors for a map, a graph or any information visualization can be one of the most satisfying but also most frustrating design experiences when trying to develop aesthetically pleasing, engaging, societally relevant, perpetually salient, and cognitively supportive visualizations.

This is because color has various perceptual properties and semantic connotations that are difficult to control for a visualization designer. The color challenge relates to any information visualization, across all of the sciences. Some color principles have been tested empirically by cartographers and information visualization experts, and some of this research has led to widely used software to support users in selecting appropriate color schemes for maps, statistical graphics, and other multivariate information visualizations.

In my talk I report on an empirical study about commonly employed color schemes in neuroimaging for brain activity maps (i.e., spectral/rainbow and heated body schemes), and in cartographic maps (i.e., univariate and bi-variate schemes). Experts in neuroimaging (N=134) and geovisualization (N=197), and also participants (N=486) sampled from the general public (via Amazon Mechanical Turk) were recruited worldwide to participate in an online web-based study. Participants were asked to interpret shown data and provide trust assessments with 1) brain maps showing various brain activity states (i.e., from normal brains to brain death), and 2) corresponding cartographic displays, depicting the outcomes of an environmental sustainability model for a particular country that matched the employed brain state descriptions. We find significant differences in data interpretations and trust ratings across color schemes and disciplinary expertise. Contrary to our hypothesis that domain experts' interpretations and ratings would be least affected by a particular color scheme in their own field, we did find that neuroimaging experts were most strongly influenced in their assessments, due to differing color schemes.

Our empirical evidence on the effect of color schemes on people's data interpretations and trust in the data further informs empirically evaluated design guidelines and metaphors to create aesthetically pleasing, engaging, societally relevant, perpetually salient, and cognitively supportive visual displays that also include multi-level graph networks of big data.

3.4 Dagstuhl Exhibit Debut: Places & Spaces: Mapping Science

Francis Harvey (Leibniz Institut für Länderkunde – Leipzig, DE) & Katy Börner (Indiana University – Bloomington, US) (Monday 19:00-20:00)

We organized an art exhibit contributed by scientists and science maps contributed by artists and used this to stimulate discussion. It is now permanently available online on the Dagstuhl Webpage: `https://www.dagstuhl.de/de/ueber-dagstuhl/kunst/places-spaces/`.

Figure 2 Pictures from the Places and Spaces: Mapping Science exhibitions.

As a part of the exhibition, we displayed contributions from participants in the seminar (maps or macroscopes), as well as a small selection from the international *Places & Spaces: Mapping Science* exhibit.

Drawing from across cultures and across scholarly disciplines, the Places & Spaces: Mapping Science traveling exhibit demonstrates the power of maps (visualizations) to address vital questions about the contours and content of scientific knowledge. As of March 2021, the exhibit features 100 framed maps, 24 macroscopes, an award-winning short film, touchscreen interactives, and sculptural elements created by more than 230 leading experts in the natural, physical, and social sciences, scientometrics, visual arts, science policy, and the humanities.

The maps have been displayed in 30 countries on six continents. Places & Spaces showcases innovative approaches to data visualization, critical for making sense of the large streams of data we confront on a daily basis. Ranging from reproductions of early maps of our planet, to the first maps showing the terrain of science, to maps showing the national mood through tweets over the course of a day, the exhibit touches on subject matter as diverse as polar bear habitat, forecasting epidemics, and the settings of Victorian poems.

In 2015 Places & Spaces expanded from exhibiting static maps of science to include interactive data visualizations we call macroscopes. Macroscopes are software tools that help one focus on patterns in data that are too large or complex to view unaided. Interactive by nature, one can use them to visually explore data and to ask and answer new questions.

Maps serve as navigational tools, documenting the landscape, warning of hazards, and highlighting potential routes of travel. Science maps chart the more abstract spaces of data and knowledge, helping forecast new fields of inquiry. Individually and as a whole, the science maps in Places & Spaces: Mapping Science use data to tell meaningful stories that both the scientist and the layperson can understand and appreciate.

Macroscopes are software tools that help us focus on patterns in data that are too large or complex to see with the naked eye. Interactive by nature, they are best used to visually explore data and to ask and answer new questions. Each macroscope featured was selected as an outstanding example of how visualization can reveal trends and patterns in data.

3.5 Knowledge Dynamics Reconstruction + Gargantext

David Chavalarias (CNRS – Paris, FR), (Tuesday 19:00-20:00)

The little prince asked Saint-Exupery to draw him a sheep, but what if he had asked him to draw Science? How could he have done it and what could we have learned from it? We address the question of "drawing science" by taking advantage of the massive digitization of scientific production, and focusing on its body of knowledge. We demonstrate how we can reconstruct, from the massive digital traces of science, a reasonably precise and concise approximation of its dynamical structures that can be grasped by the human mind and explored interactively. For this purpose, we formalize the notion of level and scale of knowledge dynamics as complex systems and we introduce a new formal definition for phylomemetic networks as dynamical reconstruction of knowledge dynamics. We propose a new reconstruction algorithm for phylomemetic networks that outperforms previous ones and demonstrate how this approach also makes it possible to define a new temporal clustering on dynamical graphs. Finally, we show in case studies that this approach produces representations of knowledge dynamics close to the ones that can be obtained by synthesizing the points of view of experts on a given domain.

3.6 Creating, Visualizing and Accessing Global Models of Science: Scopus and PubMed

Kevin W. Boyack (SciTech Strategies Inc. – Albuquerque, US), (Wednesday 19:00-20:00)

The mapping of scientific topics and fields has been increasing over time. Most science maps are what we would call "local maps," meaning that they are based on relatively small sets of literature (thousands to tens of thousands of documents) retrieved from bibliographic data providers using keyword-based or journal-based searches. In contrast, "global maps" of science are based on full databases (millions to tens of millions of documents), and these are only created by few laboratories in today's world. In this presentation we explain how global maps (also called global models) are created, give examples of the differences between local and global models and give examples of some current global models. In short, we advocate for the creation and use of global models.

We use the example of a set of models created by several teams using a dataset of Astronomy papers. Most teams used the local approach, applying different relatedness measures and clustering algorithms to the 110K+ papers. Our team looked up the positions of the Astronomy papers in our global model that was comprised of 40M+ documents from Scopus. While there were some similarities between the many models there were also many differences. From the global perspective we found that: (1) the Astronomy set was missing tens of thousands of relevant papers, and (2) that it contained thousands of papers that were weakly linked within the set and that likely didn't belong. Both of these things suggest that local models are less accurate than global models. In addition, a recently published independent study measured the relative accuracies of the different approaches to modeling the Astronomy dataset and found that the global model was far more accurate than any of the models in three of the four new metrics that were tested.

Pros and cons of local and global models were compared. The key advantages of local models are that they can be created by any researcher with institutional access to a bibliographic database using one of several different software packages. The key disadvantage of global models is that they require a significant upfront investment in data access, time and computing to create. However, once created, the global model has the advantage of covering all of science, higher accuracy, and they can be used for a multitude of analyses.

Finally, we describe two existing global models of science. First, Elsevier hosts a global model of science in their SciVal tool. It contains nearly 100,000 "topics," or clusters of documents, and is being increasingly used by different institutions to identify hot topics, very focused areas of leadership, and areas of potential collaboration among other tasks. Second, we recently created a model of PubMed that contains nearly 29,000 clusters of documents. This model is open source, is available as a Tableau workbook, and can be downloaded and used by anyone. The paper is at `https://www.nature.com/articles/s41597-020-00749-y` and the model is at `https://doi.org/10.6084/m9.figshare.12743639`.

3.7 World Processor and the Shape of Science

Ingo Günther (Worldprocessor Studio – New York, US), (Thursday 19:00-20:00)

A brief introduction into educational and aesthetically driven experiments and experiences with data representation on illuminated spheres (also known as globes). The presentation tracks 30 years of user/audience interaction with data representations on illuminated spheres (also known as globes) and explores its limits in educational, infotainment and art settings. 20 case studies ("World-Processor") illustrate the challenges of mapping on a sphere, from information overload to scaling and deciding what not to show so as to create memorable (=operational) narratives. Special consideration is given to the question of when and where to break cartographic conventions.

The presentation evaluates the entertaining and informational qualities of short, dynamic, movie-like process mapping and examines the role of audio and its potential to be enhancing or counterproductive. Deliberate breaks of cartographic conventions (exemplified in "Topography Drive") aim to shed new light and allow a fresh perspective on dimensions and proportions on a global scale. It becomes clear that working with the globe shape and a geographic mental model presents a problematic template for modeling the universe of science. We can all recognize but hardly draw the outlines of continents, let alone on a sphere. Worse, the continental geographic shapes of the "globe" – our established world model – fundamentally misrepresent the properties and weights of what constitutes the world "that is the case."

There is a need for a new vision/representation of what we identify as the totality of the world's reality. The World-Processor project endeavors to examine the constraints and advantages as well as the fallibility of the centrist universe model with respect to a possible "Shape of Science."

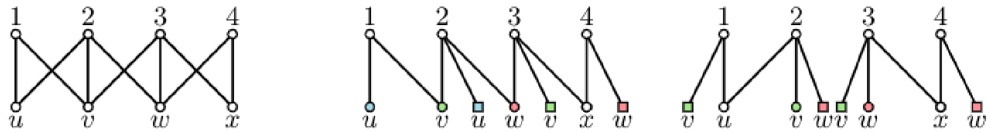

Figure 3 (a) A graph instance G; (b) an optimal solution with three splits, involving three different vertices; (c) an optimal solution with two split vertices (but with more vertices in total).

4 Working Groups

4.1 Working Group 1: "Splitting Vertices in 2-Layer Graph Drawings"

Reyan Ahmed (University of Arizona, Tucson, AZ, USA), Patrizio Angelini (John Cabot University, Rome, Italy), Giuseppe Di Battista (Roma Tre University, Rome, Italy), Philipp Kindermann (Universität Trier, Trier, DE), Karsten Klein (University of Konstanz, DE), Stephen Kobourov (University of Arizona, Tucson, AZ, USA), Martin Nöllenburg (TU Wien, Vienna, Austria), Antonios Symvonis (NTUA, Athens, Greece), Markus Wallinger (TU Wien, Vienna)

2-layer networks model the relationships between two disjoint sets of entities in various applications. The inter-layer edges are naturally modeled as a bipartite graph and visualized with a 2-layer drawing with vertices placed on two parallel lines. Such 2-layer drawings are widely used to demonstrate the relationship between two partitions. Minimizing the number of edge crossings can dramatically improve the readability of such visualizations. Crossing minimization in 2-layer graph drawings is a prominent problem, but obtaining a drawing with no or just a few crossings is often an illusive goal. We propose to use vertex splitting as an operation to reduce the number of crossings by replacing selected vertices on one layer by two (or more) copies and suitably distributing their incident edges among the copies. We study the algorithmic problem of minimizing the number of splitting operations or of split vertices, or of the maximum number of splits per vertex, to obtain a crossing-free drawing. We prove NP-hardness results when both layers are permutable, and give linear-time algorithms when one layer is fixed in the input.

The *vertex-split* operation (or *split*, for simplicity) for a vertex v deletes v from G, adds two new copies v_1 and v_2 (in the original vertex subset of G), and distributes the edges originally incident to v among the two new vertices v_1 and v_2. Placing v_1 and v_2 independently in the 2-layer drawing can reduce the number of crossings; see Fig. 3. Vertex splitting has been studied in the context of the *splitting number*, which is the smallest number of vertex-splits needed to obtain a planar graph. The splitting number problem is NP-complete, even for cubic graphs [1], but the splitting numbers of complete and complete bipartite graphs are known [2, 4].

We study variations of the algorithmic problem of constructing planar 2-layer drawings with vertex splitting. In visualizing graphs defined on anatomical structures and cell types in the human body [3], the two vertex sets of G play different roles and vertex splitting is permitted only on one side of the layout. This motivates our interest in splitting only the bottom vertices. The top vertices may either be specified with a given context-dependent input ordering, e.g., alphabetically or according to some importance measure, or we may be allowed to arbitrarily permute them to perform fewer vertex splits. Also motivated from the above application, we consider a setting in which crossings are allowed but we aim for compact neighborhood ranges. In this setting, the goal is to minimize the maximum *span* of the vertices on the top layer, i.e., the maximum distance between a first and a last neighbor in the order of the bottom vertices, given a budget of k vertex-splits on the bottom layer.

We prove that for a given integer k it is NP-complete to decide whether G admits a planar 2-layer drawing with an arbitrary permutation on the top layer and at most k vertex splits on the bottom layer. NP-completeness also holds if at most k vertices can be split, but each an arbitrary number of times. If, however, the vertex order of V_t is given, then we can compute a planar 2-layer drawing in linear time, both for minimizing the total number of splits and for minimizing the number of split vertices.

References

1. L. Faria, C. M. H. de Figueiredo, and C. F. X. Mendonca. Splitting number is NP-complete. *DAM*, 108(1):65–83, 2001.
2. Nora Hartsfield, Brad Jackson, and Gerhard Ringel. The splitting number of the complete graph. *Graphs and Combinatorics*, 1(1):311–329, 1985.
3. Hubmapconsortium. `https://hubmapconsortium.github.io/ccf-asct-reporter/`.
4. B. Jackson and G. Ringel. The splitting number of complete bipartite graphs. *Archiv der Mathematik*, 42(2):178–184, 1984.

4.2 Working Group 2: "User Studies for Science Mapping"

Angela Zoss (Duke University, US), Andrea Scharnhorst (KNAW, The Netherlands), Lynda Hardman (Centrum Wiskunde & Informatica, Netherlands), David Chavalarias (CNRS, France), Jean-Daniel Fekete (INRIA, France), Sara Fabrikant (University of Zurich, Switzerland), Thomas Koehler (TU Dresden, DE)

During the workshop, the group brainstormed a few ideas for projects related to user studies. Both Andrea (comparison of the use of the open source Dataverse platform) and Lynda (user study comparing 2D representations to 3D representations) have local projects that would benefit from user testing. All participants were convinced that despite the existence of evaluation studies already, there is still a great need to understand better how visualizations in general and large-scale maps in particular are used in navigating existing and producing new knowledge. For example, in other discussions, attendees posed the idea of developing a list of standard use cases that could be applied to a variety of science maps. In the user studies group, Jean-Daniel raised the idea of benchmarking, or developing success metrics like precision and recall that could then be used to evaluate science maps. He proposed running a contest to see how well maps can meet these benchmarks. A lively discussion, largely between Jean-Daniel and David, explored the idea of success for science maps and how users play a role in science map evaluation. In contrast to the idea of universal benchmarking for maps, David's Gargantext system is organized around the idea of a science map that is highly customized for a particular user or group exploring a specific question. The success for this system is in the subjective user experience, rather than testing the system against a set of objectively "correct" results.

Eventually, a core group (Angela Zoss, Andrea Scharnhorst, Lynda Hardman, David Chavalarias, Sara Fabrikant) started to explore a concrete idea. Rather than resolve the long-standing debate about the balance between objective and subjective evaluation criteria, the group coalesced around the need to create a survey of science mapping evaluation techniques. Similar to recent work in the information visualization community, this survey would constitute a broad scan of literature on science mapping and related fields and catalogue types of user studies that have been applied to science maps. The group emphasized the need to search across disciplinary silos, as there are many related fields of study that don't use the term "science mapping." The group plans to use David's Gargantext system to compile relevant literature, iteratively improving search terms to identify as many papers as possible. The plan is to categorize the science maps and the related user studies. Finally, the group will reflect on the difficulties of applying different user study methodologies, including the obstacles researchers face when attempting to translate methods from one field to another.

The group started with a first literature search during the Dagstuhl workshop, met once again afterwards, and participated at the Dagstuhl Re-connect meetings in May and June. We produced various queries and related mappings of the relevant literature. The work has been paused due to a new release of the Gargantext system, which better enables collaborative working. But, in principle as was discussed on the Re-connect meetings would be relevant for either the IEEE special issue or another publication venue. Work is expected to become resumed in September again.

The group plans to submit a paper "How to evaluate science maps/map-like visualizations, and what can we learn from each other's evaluation practices?" Visualizations are part of the epistemological, methodological instrumentation all academic fields use. They are also an object of research themselves in those fields developing new methods for visualizations. There is obviously an overlap between the "making" of visualizations and the "use of visualizations." This paper zooms into a specific class of visualizations: science maps. By science maps we mean mapping out literature/documented kind of information (up to the very end of ideas). We use science mapping techniques to map out the scientific literature around science maps. So, there is a kind of second order operation involved. In particular, we are starting with a bibliometric approach, to identify: (a) baseline collection and (b) show the disparity, but also concentration of discourses around evaluation practices in the InfoVis area. We develop a higher-level ordering system (matrix) to display, discuss and compare "evaluation dimensions." The paper itself is based on an iterative process (which we transparently reflect about) from data collection, mapping, zooming in and close reading, and adapting the data collection. We restrict the close reading part and the results coming out of that to some cells/cases of the overall matrix/ordering system.

4.3 Working Group 3: "Human Reference Map: Mapping the Human Body at Single-Cell Resolution"

Katy Börner (Indiana University, US), Ingo Günther (New York, US), Francis Harvey (Leipzig University, DE), Michael Kaufmann (University of Tübingen, DE), Alexander Wolff (University of Würzburg, DE)

The NIH-funded HuBMAP project aims to develop a common coordinate framework (CCF) for the healthy human body, https://portal.hubmapconsortium.org. The framework needs to capture many scales–from the body scale to organs to single cells; see Fig. 4. It must support cataloging different types of individual cells, understanding the functions of and relationships between those cell types, and modeling their individual and collective function.

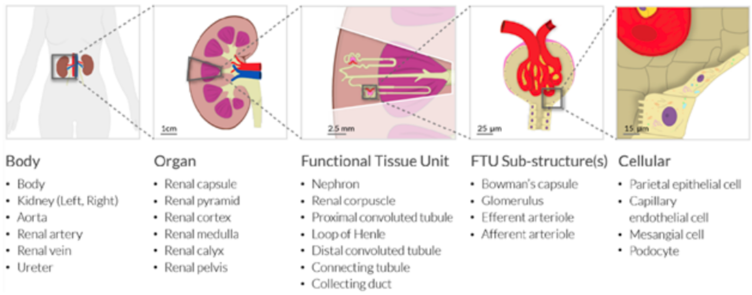

Figure 4 Multi-level spatial reference system for anatomical structures.

The relationships of cells inside of organs inside of the human body as well as the biomarkers used to *characterize* different cell types are captured in so called ASCT+B tables. The anatomical structures (AS) partonomies and information for the cell types (CT) each AS contains can be explored at `https://hubmapconsortium.github.io/ccf-asct-reporter`.

The vasculature interlinks the different organs currently captured in ASCT+B tables, see Fig. 5. We are interested to interlink the vasculature branching tree with the AS part-of tree for kidney, heart, and other organs using algorithms similar to those shown in Fig. 6.

Figure 5 Exemplary vasculature tree layout.

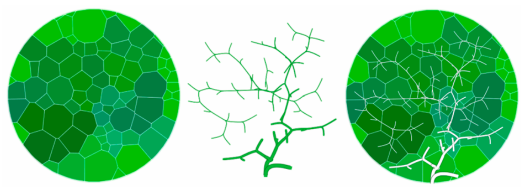

Figure 6 An example of a Voronoi diagram and flowmap on the order of Passeriformes (perching birds), by Horn *et al.* 2009.

Participants

- Katy Börner
Indiana University – Bloomington, US
- Ingo Günther
Worldprocessor Studio – New York, US
- Francis Harvey
Leibniz Institut für Länderkunde – Leipzig, DE
- Michael Kaufmann
Universität Tübingen, DE
- Alexander Wolff
Universität Würzburg, DE

Remote Participants

- Patrizio Angelini
John Cabot University – Rome, IT
- Michael A. Bekos
Universität Tübingen, DE
- Kevin Boyack
SciTech Strategies Inc. – Albuquerque, US
- David Chavalarias
CNRS – Paris, FR
- Giuseppe Di Battista
University of Rome III, IT
- Sara Irina Fabrikant
Universität Zürich, CH
- Jean-Daniel Fekete
INRIA Saclay – Orsay, FR
- Lynda Hardman
CWI – Amsterdam, NL & Utrecht University, NL
- Ben Jacobsen
University of Arizona – Tucson, US
- Philipp Kindermann
Universität Trier, DE
- Karsten Klein
Universität Konstanz, DE
- Stephen G. Kobourov
University of Arizona – Tucson, US
- Thomas Köhler
TU Dresden, DE
- Vincent Larivière
University of Montreal, CA
- Tamara Mchedlidze
Utrecht University, NL
- Guy Melançon
University of Bordeaux, FR
- Staša Milojevic
Indiana University – Bloomington, US
- Filipi Nascimento Silva
Indiana University – Bloomington, US
- Martin Nöllenburg
TU Wien, AT
- Adam Ploszaj
University of Warsaw, PL
- Sergey Pupyrev
Facebook – Menlo Park, US
- Chrysanthi Raftopoulou
National Technical University of Athens, GR
- Andrea Scharnhorst
Royal Netherlands Academy of Arts and Sciences, NL
- André Skupin
San Diego State University, US
- Cassidy Rose Sugimoto
Indiana University – Bloomington, US
- Antonios Symvonis
National TU – Athens, GR
- Markus Wallinger
TU Wien, AT
- Angela Zoss
Duke University – Durham, US