Report from Dagstuhl Seminar 21401

Visualization of Biological Data – From Analysis to Communication

Edited by

Karsten Klein[1], Georgeta Elisabeta Marai[2], Kay Katja Nieselt[3], and Blaz Zupan[4]

1 **Universität Konstanz, DE,** karsten.klein@uni-konstanz.de
2 **University of Illinois Chicago, US,** gmarai@uic.edu
3 **Universität Tübingen, DE,** kay.nieselt@uni-tuebingen.de
4 **University of Ljubljana, SI,** blaz.zupan@fri.uni-lj.si

Abstract

Technological advancements in biology allow us to collect and generate a large quantity of data and pose a significant challenge to data interpretation and understanding. Addressing this challenge requires a blend of methodology from data visualization, bioinformatics, and biology. This methodology encompasses perception and design knowledge, algorithm design, techniques for analyzing and visualizing big data, statistical approaches, and specific domain knowledge for different application problems. In particular, it is essential to develop robust and integrative visualization methods combined with computational analytical techniques and approaches to communicate the outcomes visually. The purpose of Dagstuhl Seminar 21401, "Visualization of Biological Data – From Analysis to Communication," was to bring together researchers from various fields to discuss the state of the art, to debate means of advancing science in the field of visualization of biological data, and to foster the development of our international community.

Seminar October 3–8, 2021 – http://www.dagstuhl.de/21401
2012 ACM Subject Classification Applied computing → Bioinformatics; Applied computing → Life and medical sciences; Human-centered computing → Visualization
Keywords and phrases Bioinformatics, biology, Imaging, interdisciplinarity, Omics, Visual analytics, visualization
Digital Object Identifier 10.4230/DagRep.11.9.1
Edited in cooperation with Beauxis-Aussalet, Emma

1 Executive Summary

Kay Katja Nieselt (Universität Tübingen, DE)
Karsten Klein (Universität Konstanz, DE)
Georgeta Elisabeta Marai (University of Illinois Chicago, US)
Blaz Zupan (University of Ljubljana, SI)

Advances in technology have turned biology into data-driven research. High-throughput and high-resolution techniques help us generate and collect vast amounts of data to be explored, analyzed, and turned into knowledge or actionable models. This abundance of biological data creates a substantial challenge in processing, analysis, and modeling. A popular way to address this challenge is through visual representation and analysis of the data.

Creating compelling visualizations of biological data requires combining data visualization, bioinformatics, statistics, and computational biology. The field of biological visualization is interdisciplinary and involves collaboration between researchers from different areas.

Our aim with this Dagstuhl Seminar was to bring together researchers from multiple disciplines to discuss how to continue the interdisciplinary dialogue and foster the development of an international community concerned with biological visualization. We aimed to examine the state of the art and find areas to advance the research that might benefit from the joint efforts of all groups involved.

Our initial aim, expressed in the seminar proposal, was to explore the following four topics:

- data abstraction to support building custom visual tools of biological data,
- interactive analysis for biological data exploration,
- collaboration and communication through new tools,
- curriculum for teaching visualization in bioinformatics.

We have discussed these topics in the first two days of our five-day seminar and gradually came out with the following six working groups:

- facilitating cross-expertise exploration in explainable AI for multi-omics via visualization,
- visions for the lab notebook of the future,
- visual analytics of multilayer networks representing knowledge graphs,
- recommendations for designing visual, interpretable, and deep learning-based analytics pipelines in medical imaging,
- semantically enabled biomedical cartooming,
- a curriculum for the future of biological visualization.

Notice that with this new set of topics for the working groups, our seminar still closely followed our initial aim to explore the space of data abstractions, interactive analysis, and design of tools to support collaborations.

We have developed the schedule for the seminar based on our experience and expertise in previous successful Dagstuhl seminars. We aim to emphasize the balance between prepared talks and panels and breakout groups for less structured discussions focused on a selection of highly relevant topics. Three types of plenary presentations were available to participants who had indicated an interest in presenting during the seminar:

- Overview talks (20 minutes plus 10 minutes for questions)
- Regular talks (10 minutes plus 5 minutes for questions)
- Panel presentations (5 minutes per speaker followed by a 20 – 25 minute discussion)

The breakout groups met multiple times for several hours during the week and reported to the overall group on several occasions. This format successfully brought bioinformatics and visualization researchers onto the same platform and enabled researchers to reach a shared, deep understanding through their questions and answers. It also stimulated fruitful discussions that all participants deeply appreciated.

We have organized the seminar during the COVID-19 pandemic. Due to various regulations and quarantines, slightly above half of the participants attended in person, while the other participants attended online. The meeting took the hybrid form, and we thank Dagstuhl for equipping the seminar rooms with suitable hardware. Still, we found the organization of the hybrid meeting challenging, to say the least, as it imposed constraints on the discussion and engagement of everyone. An all-important part of Dagstuhl's experience is off-line meetings during meals, trips, or long evenings, which online participants miss. Thus, also following the responses in the participants' survey, we would endorse their recommendation that Dagstuhl should return to the previous, non-hybrid format of seminars once the pandemic stops.

This report describes in detail the outcomes of our meeting. At the present stage, the outcome includes a set of white papers summarizing the breakout sessions, overviews of the talks, and an emerging detailed curriculum for future biological and medical visualization eduction.

Acknowledgements

We would like to thank all participants of the seminar for their contributions and lively discussions; we also would like to thank the scientific directorate of Dagstuhl for providing us with the opportunity to organize this seminar. Finally, the seminar would not have been possible without the untiring help of the (scientific) staff of Dagstuhl, including Ms. Jutka Gasiorowski and Ms. Susanne Bach-Bernhard.

2 Table of Contents

3 Overview of Talks

3.1 Embracing the complexity of reality

Jan Aerts (Hasselt University – Diepenbeek, BE)

In this talk, we briefly go over what explainable AI could mean for data analysis, and explain how our own approach of topological data analysis (TDA) can help with that endeavor by providing contextual information. We illustrate this through two use cases: analysis of polysomnography data, and supporting patient/clinician discussion on cancer treatment.

3.2 Collaboration and Communication in Science (Using Data Visualization)

Andreas Bueckle (Indiana University – Bloomington, US) and Katy Börner (Indiana University – Bloomington, US)

As science becomes more complex, communication across interdisciplinary teams becomes ever more essential to solve complex problems arising in disease and health, the physical laws of nature, and the functioning of human societies. In this talk, we present three ongoing efforts that aim to empower researchers and domain experts to communicate and collaborate on the challenges of our time: First, we present a paper on visualizing big science projects (`https://doi.org/10.1038/s42254-021-00374-7`) using a dataset of 13,893 publications and 1,139 grants by 21,945 authors cited more than 333,722 times. The work aims to communicate the global and interdisciplinary reach and impact of big science projects and their evolution via distinct project phases. Second, we present user interfaces developed for the Human Biomolecular Atlas Program (HuBMAP), a multi-year, multi-million, international effort funded by the National Institutes of Health (NIH) with the goal "to develop an open and global platform to map healthy cells in the human body" (`https://commonfund.nih.gov/hubmap`). Specifically, we show how the Registration User Interface (`https://hubmapconsortium.github.io/ccf-ui/rui/`) and the Exploration User Interface (`https://portal.hubmapconsortium.org/ccf-eui`) have been developed as web-deployed, 3D tools for users to register and browse through human tissue blocks from more than a dozen organs in their spatial and semantic context. Finally, we discuss the Scalable Precision Medicine Knowledge Engine (SPOKE) Visualizer, which allows users to explore a complex network of 3 million nodes and 30 million edges involving food, genes, proteins, diseases, and symptoms, among others through a series of network and map visualizations (`https://cns-iu.github.io/spoke-vis`). The site was designed for novice users interested to understand the coverage and quality of SPOKE data and expert users interested to analyze and optimize the interlinked knowledge graphs in SPOKE.

3.3 Explainable AI – On the importance of domain awareness for the creation of visual feedback on black box classifier decisions on imaging data

Katja Bühler (VRVis – Wien, AT)

Joint work of Katja Bühler, David Major, Dimitrios Lenis, Maria Wimmer, Astrid Berg

Recent research in AI addresses architectures and network types providing a certain level of interpretability by design. This is opening up new avenues for future research on how this information can be employed for creating visual feedback to create trust and provide transparency on the decision and its reliability to end users. However, in particular for imaging data, many of these novel methods are not yet widely applied to clinical imaging data in a real-world context.

On the other hand, classical "black box" classifiers for imaging data based on ANNs are well established, but do not provide any insight into the reliability and the reasons behind the result. This is preventing in particular a straight forward solution to visually communicate the causality between input data and the classifiers' decision to end user. Existing methods take different strategies to attribute salient regions but generally fail either in creating focused and intelligible feedback tailored to the specific task or come with the danger to operate out of domain, making the classifiers decision, and consequently also the visualization building on it, unreliable. I will highlight the importance of domain awareness in designing task specific and trustworthy visual feedback and present a novel method for trustworthy XAI for pathology classification on medical images based on two of our recent papers [1, 2]

References

1 Major D., Lenis D., Wimmer M., Sluiter G., Berg A., and Bühler, K. (2020), *Interpreting medical image classifiers by optimization based counterfactual impact analysis*, IEEE International Symposium on Biomedical Imaging, Iowa City, 2020.

2 Lenis D., Major D., Wimmer M., Berg A., Sluiter G., Bühler K. (2020) *Domain Aware Medical Image Classifier Interpretation by Counterfactual Impact Analysis.*. In: Martel A.L. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science, vol 12261. Springer, Cham.

3.4 Explainable and Interactive AI: Overview and Open Research Challenges

Mennatallah El-Assady (ETH Zürich, CH)

URL https://explainer.ai/

Interactive and explainable machine learning can be regarded as a process encompassing three high-level stages:

1. Understanding machine learning models and data,
2. Diagnosing model limitations using eXplainable AI (XAI) methods, and
3. Refining and optimizing models interactively.

In this talk, I review the current state-of-the-art of visualization and visual analytics techniques by grouping them into these three stages. In addition, I argue for expanding our approach to explainability by adapting concepts, such as metaphorical narratives, verbalization, and gamification. I further introduce the explAIner.ai framework for structuring the process of explainable artificial intelligence and interactive machine learning, and operationalizing it through a TensorBoard plugin. Lastly, to derive a robust XAI methodology, I present some first steps to extract XAI strategies and mediums by transferring knowledge and best practices from other disciplines.

3.5 Impact of Model Prediction on Human Decision Making Confidence

Carsten Görg (University of Colorado – Aurora, US)

We conducted a pilot study to assess the impact of introducing model prediction in an existing clinical workflow for classifying Adamantinomatous Craniopharyngioma (ACP). We designed an experiment with 3 conditions: (1) viewing images only, (2) viewing images plus a model prediction, and (3) viewing images providing a model prediction and using the What-If-Tool (WIT) for exploring conterfactuals. In each condition we used 28 cases, about half ACP and non-ACP. Most cases had either an MRI or a CT image, some cases had both. We enrolled 2 expert subjects, a neurosurgeon and a neuroradiologist. We collected measures of performance, confidence and difficulty. Performance was similar across conditions but different between subjects. The confidence and difficult measures show the expected pattern of high confidence for not difficult cases, but also less confidence for non-ACP cases. Both experts found the similarity feature in WIT useful, but the tool itself too difficult to use. Limitations include the small number of subjects and limited scope of the dataset.

3.6 Designing a Topic-Based Literature Exploration Tool – A Neuroscience Exploratory Study

Lynda Hardman (CWI – Amsterdam, NL)

Exploring an extensive body of literature can be facilitated through topic-based, rather than article-based, interaction. Neuroscience researchers are interested in exploring relations between topics, such as anatomical regions of the brain and diseases that affect them. Given the three-dimensional nature of the brain, we postulate that supporting the exploration of relations between neuroscience topics in Augmented Reality could improve and extend existing literature exploration workflows. We identify visualization and interaction design requirements for distant reading of neuroscience literature using a user-centered approach. Using an existing concept analysis of tens of thousands of neuroscience papers, we designed an Augmented Reality environment to support distant reading of relations between brain regions and brain diseases.

3.7 Abstraction for Bioviz

Lawrence Hunter (University of Colorado – Aurora, US)

Abstractions convey groupings of facts as a unit, by omitting "details". To express knowledge, a system must select which "facts" to present. This selection is based both on the particular task the visualization is to support, and on the shared understanding of scientists. One useful way to make progress is with formal representations of biomedical knowledge. These formal representations are often in the form of knowledge graphs where the nodes (and sometimes arcs) are drawn from computational ontologies. These Knowledge Graphs (KGs) can be either tools to visualize data structured by the graph, or to visualize subgraphs that are relevant to particular biomedical tasks. One possible concrete challenge would be tools for the generation of molecular biology "cartoon" summary figures.

3.8 From Graphs and Hypergraphs to Maps and MetroMaps

Stephen G. Kobourov (University of Arizona – Tucson, US)

Relational datasets are often modeled with graphs and hypergraphs: objects become vertices and relationships become edges and hyperedges. Algorithms for graph and hypergraph visualization aim to represent such data in an effective and aesthetically pleasing way. From a theoretical point of view, the underlying problems give rise to algorithmic and complexity questions. From a practical point of view, building functional visualization systems is associated with questions of scalability and usability. GMap and MetroMaps are general visualization frameworks for utilizing familiar data representations and metaphors, such as geographic maps and metro maps. Such representations are more intuitive, as people are already proficient with such maps and standard interactions via panning and zooming.

3.9 Reproducibility and Reusability in Interactive Visual Analytics

Alexander Lex (University of Utah – Salt Lake City, US)

Interactive analysis is an important part of the data science process. It enables analysts to directly interact with the data, exploring it with minimal effort. Unlike code, however, an interactive visualization session is ephemeral and cannot be easily shared, revisited or reused. In this talk, I will sketch approaches to "Literate Visual Analysis". I will show how we can leverage provenance data of an analysis session to create well-documented and annotated visualization stories that enable reproducibility and sharing. I will also introduce work on inferring analysis goals, which allows us to understand the analysis process at a higher level. Understanding analysis goals enables us to enhance interactive capabilities and even re-use visual analysis processes. I will conclude by demonstrating how this provenance data can be leveraged to bridge the gap between computational and interactive environments.

3.10 Collaboration in a Test of Time Project

Georgeta Elisabeta Marai (University of Illinois Chicago, US)

RuleBender is an open-source system for the integrated visualization, modeling and simulation of rule-based intracellular biochemistry. RuleBender has been remarkably stable over the past ten years, and has been used by at least hundreds of computational biologists to do real and high impact research. It has been adopted at over 40 institutions as a research and educational tool. We reflect on the lessons learned from the design, development, and deployment of this successful system. We particularly emphasize the activity-centered design paradigm and the close interaction with domain experts that allowed RuleBender to better serve the needs of the systems biology community.

3.11 Proteomics Data – the Next Generation

Lennart Martens (Ghent University, BE)

This talk was derived from an ad-hoc question posed to the Seminar Participants regarding any input they might have on the visualisation (and analysis) challenges of a large amount of newly generated proteomics results, based on a proteome-wide analysis of post-translational modifications derived from a large and heterogeneous set of public proteomics data for human and mouse samples.

Computational proteomics has evolved extensively over the past decade, introducing the first widely successful machine learning approach with the Percolator algorithm for post-processing of identification data [1], and slowly expanding the capabilities of predictive algorithms for analyte behaviour during liquid chromatography separation and fragmentation. Despite clear advances in these areas, however, the adoption of such predictions in identification pipelines was very slow to non-existent. This was primarily due to the already quite capable traditional identification algorithms, especially when complemented with Percolator post-processing.

However, renewed interest in the use of the prediction of analyte behaviour was created by the field's diversification into more complex analyses such as metaproteomics, proteogenomics, immunopeptidomics, and open modification searches, all of which revealed inherent limitations in existing, traditional search engines [2].

As it turns out, machine learning based predictions are highly effective at solving the issues encountered in such complex DDA proteomics approaches, and the production of highly performant algorithms has soared as a consequence.

Interestingly, the availability of vast amounts of public data have also enabled a breakthrough in the types of machine learning algorithms that can now be employed on proteomics data, which has seen a surge in the application of complex neural networks (so-called deep learning approaches) in the field. When provided with enough data, these deep learning algorithms deliver extremely good predictions, which will in turn fuel a more sensitive and more robust interpretation of already acquired data.

In this lecture, our latest, cutting-edge developments in machine learning based algorithms for computational proteomics have therefore been briefly presented, alongside our applications of these algorithms to the complete re-analysis of all publicly available human and mouse proteomics data.

Moreover, the lecture shows our first efforts at making the results from these data widely available, which takes two main forms: the Scop3P system to show modifications in context on linear sequence and 3D structure [3], and the Tabloid Proteome approach to derive protein association networks from co-occurring proteins across different experiments [4, 5].

The resulting discussion proved highly interesting and informative, including the possibility of including the results set as a challenge in the 2022 BioVis Conference, the analysis of the results using models originally derived for linguistics work, and the application of upset visualisations and interactive histograms.

References

1 Lukas Käll„ Jesse D. Canterbury, Jason Weston, William Stafford Noble, and Michael J. MacCoss. *Semi-supervised learning for peptide identification from shotgun proteomics datasets*. Nature methods 4, no.11, pp.923–925, 2007.

2 Niklaas Colaert, Sven Degroeve, Kenny Helsens, and Lennart Martens. *Analysis of the resolution limitations of peptide identification algorithms*. Journal of proteome research 10, no.12, pp.5555–5561, 2011.

3 Pathmanaban Ramasamy, Demet Turan, Natalia Tichshenko, Niels Hulstaert, Elien Vandermarliere, Wim Vranken, and Lennart Martens. *Scop3P: a comprehensive resource of human phosphosites within their full context*. Journal of Proteome Research 19, no. 8, pp.3478–3486, 2020.

4 Surya Gupta, Kenneth Verheggen, Jan Tavernier, and Lennart Martens. *Unbiased protein association study on the public human proteome reveals biological connections between co-occurring protein pairs*. Journal of proteome research 16, no.6, pp.2204–2212, 2017.

5 Surya Gupta, Demet Turan, Jan Tavernier, and Lennart Martens. *The online Tabloid Proteome: an annotated database of protein associations*. Nucleic acids research 46, no.1, pp.581–585, 2018.

3.12 Bio+Med+Vis Education

Torsten Möller (Universität Wien, AT)

Joint work of Torsten Möller, Johanna Beyer, Jan Byska, Ingrid Hotz, Bara Kozlikova, Renata Raidon, Noeska Smit, Hsiang-Yun Wu

URL https://biomedvis.github.io/

In this talk we are detailing our efforts toward a curriculum for a Bio+Med+Vis course and program. In the context of the 2018 BioVis Dagstuhl and the 2020 Shonan Meeting, a working group has formed that developed the topic matrix further, and held a first Spring School to fill the materials which are now accessible to everyone for free online. We would like to brainstorm on how to continue this effort and solicit feedback from the BioVis 2021 Dagstuhl participants.

3.13 Teach (from) Your Users

Bruno Pinaud (University of Bordeaux, FR)

As a visualization researcher, I agree with Jeffery Heer: "The ultimate subject of the visualization research community is people not pictures" (Eurovis, 2019). I would like to present some lessons learned as a Principal Investigator (PI) for a multidisciplinary project, collaborating with domain experts in biological and digital humanities along with computer science colleagues. It is a delicate work to collaborate with experts you do not know, and who do not know what you can bring to them. They need money to produce data, while I need data to visualize. That is why such project unfortunately starts around its end.

A challenge, thus, is to make data experts tell you what they want as early as possible in the project, to avoid having yet another unused tool at the end of the project. So, first of all each side has to teach the other a bit about its domain. For instance, in visualization, we have mantras. They are good starting points, especially "Details first, show context, overview last" [1]. To conclude, never forget the golden rule: "Keep it simple".

References

1 Timothy Luciani, Andrew Burks, Cassiano Sugiyama, Jonathan Komperda, and G. Elisabeta Marai. *Details-first, show context, overview last: supporting exploration of viscous fingers in large-scale ensemble simulations*. IEEE Transactions on Visualization and Computer Graphics. 25(1), pp.1225–1235, 2019.

3.14 Exigent Visualization: When Good Enough is Better than Better

William Ray (Ohio State University – Columbus, US)

The Vis community usually talks about the dichotomy between Exploratory Data Visualization and Explanatory Data Visualization, and the contrast between the data-representation needs for discovering new things in data versus communicating something known in the data. Critically, we focus on optimizing completeness and correctness of data communication in these different modalities.

As a result of a clinical collaboration, we have recently encountered an interactive data visualization task (estimation of burn size in burn trauma) that does not fit either of these categories. Moreover, several decades of attempts to optimize the existing representations for completeness and correctness has resulted in the actual results in practice, getting worse rather than better.

A careful categorization of the different types of error that creep into the results from this visualization task, suggests that the practical results can be optimized by discarding the canonical optimizations to the visualization tool. In fact, making the tool fundamentally much less accurate, significantly decreases downstream errors by clinicians and can result in improved care.

We propose that a similar analysis of the different contributions to the total error observed in other visualization tasks may lead to similar counterintuitive opportunities to improve results.

3.15 Collaboration in BioMedVis

Timo Ropinski (Universität Ulm, DE)

Establishing fruitful collaborations in the area of BioMedVis, where visualization researchers and domain experts work towards a common goal, can be a challenging task. Besides the actual topics to be addressed, also the mechanics of collaboration need to be determined. While the prime goal is often to work on a common project, involving one or more full time researchers, this setup can seldom be obtained from the beginning. It is rather beneficial to initiate collaborations by starting with student projects of different levels. Furthermore, a balance between "give and take" must be determined.

3.16 Interactive exploratory analysis with iSEE

Charlotte Soneson (FMI – Basel, CH)

The iSEE (interactive SummarizedExperiment Explorer) R/Bioconductor software package (`https://bioconductor.org/packages/iSEE`), built on the R/shiny framework, provides a general-purpose graphical interface for exploring any rectangular dataset with additional sample and feature annotations, for example single-cell RNA-seq data. Users can create, configure, and interact with the iSEE interface, enabling quick iterations of data visualization. This facilitates generation of new scientific hypotheses and insights into biological phenomena, and empowers a wide range of researchers to explore their data in depth. iSEE also guarantees the reproducibility of the analysis, by reporting the code generating all the output elements as well as the layout and configuration of the user interface. The combination of interactivity and reproducibility makes iSEE an ideal candidate to bridge and complement the expertise of researchers, who are able to design flexible, accessible, and robust dashboards that can also be directly shared and deployed in collaborative contexts – connecting large data collections to broad audiences, thus further increasing the value of generated research data.

3.17 The process of designing, interpreting, and registering complex transcriptomic biomarkers for clinical oncology

Miha Stajdohar (Genialis – Houston, US)

We have modeled four distinct phenotypic subtypes of cancer patients. These subtypes represent the interplay of key biological processes in the tumor microenvironment and are clinically actionable. This is the first complex transcriptomic biomarker that is currently being registered with the FDA (USA) as a clinical trial assay. It will guide patient enrollment in the upcoming clinical trials, and the development of a companion diagnostic test (CDx) is underway. In this presentation, we will present the data science of design and validation of the classifier, visualization and interpretation of its predictions, and discuss the hurdles that we had to overcome to make our way into the clinical decision making.

3.18 The Code Monkey and The Parachuter: Reflecting on the Values of Interdisciplinary Collaboration in Visualization

Danielle Szafir (University of North Carolina at Chapel Hill, US)

Visualization prides itself on being a highly interdisciplinary field. We draw on techniques from a wide range of disciplines to design, implement, and evaluate visualizations. However, all too often, we find ourselves in collaborations that fail to truly contribute to multiple disciplines: we either build something exclusively for the domain (acting as the code monkey) or we stop contributing once the core visualization research question is solved (acting as the parachuter). I will reflect on contributions from my past work to explore the types of bidirectional contributions that interdisciplinary work can foster (or fail to foster) and examine how we might consider a range of ways visualization can benefit and be benefitted by cross-disciplinary collaboration.

3.19 What do visualizations explain with/for Explainable AI

Cagatay Turkay (University of Warwick – Coventry, GB)

This talk will start by exploring what explainability in AI means and why it matters. We will then cover some of the roles that visualization can play as a facilitator and enabler for explainability, as such roles emerge within the broader visualization and AI literature. As well, we will make a case for the human-centered nature of visualization, and how that is critical for thinking about explainability. The talk will then focus on the intersection between Explainable AI and Biological Data Visualization and explore advances, successes, failures, challenges and opportunities in this space.

3.20 XAI through a combination of interactions and workflows

Blaz Zupan (University of Ljubljana, SI)

URL https://orangedatamining.com

Visual analytics, a technique that combines exploratory data analysis and interactive visualization, can be an excellent tool for explainable AI (XAI). Consider single cell gene expression data, and a set of marker genes, for example. We can visualize these cells in t-SNE, expose those that express marker genes, allow users to select a cluster, perform differential expression analysis, and do Gene Ontology (GO) enrichment for explanation. The challenge is to generalize such pipelines to any biomedical dataset, and find the minimal set of components to implement it.

4 Working groups

4.1 Visions for the Lab Notebook of the Future

Jan Aerts (Hasselt University – Diepenbeek, BE), Jian Chen (Ohio State University – Columbus, US), Mennatallah El-Assady (ETH Zürich, CH), Alexander Koch (BioLizard – Gent, BE), Alexander Lex (University of Utah – Salt Lake City, US), James Procter (University of Dundee, GB), William Ray (Ohio State University – Columbus, US), Charlotte Soneson (FMI – Basel, CH), Granger Sutton (J. Craig Venter Institute – Rockville, US), Danielle Szafir (University of North Carolina at Chapel Hill, US), Cagatay Turkay (University of Warwick – Coventry, GB), and Blaz Zupan (University of Ljubljana, SI)

Our group worked on the concept of provenance in interactive data visualizations and insight generation. We kicked off our discussions with a roundtable where everyone introduced themselves and explained why they were interested in this topic. The participants' interests and motivations ranged from high-level issues and abstract methods to practical applications:

- Interactive data visualizations are under-respected in the field of computer science and misunderstood outside of the vis community.
- How can we communicate how and why we created a particular visualization?
- How do users, such as, for example, patients and doctors, interact with the data visualization tools we create?
- What are the best practices for creating interactive visualizations? Can we compile a list of recommendations?
- How can we make interactive tools more reproducible?
- How do we build tools that work for/with people instead of replacing them (e.g., computer in the loop rather than human in the loop)?

The first point of discussion was whether we should keep the discussion biovis-specific or make it more general. We decided to keep the discussion (and any follow-up efforts) general but apply the ensuing general principles and practical implementations to biological data visualization.

The basis of our group's further discussions was the lack of provenance in interactive data visualizations. We formulated a list of relevant questions around interactive bio(medical) visualizations on which we could build during our brainstorming sessions. These included:

- How do users, such as patients and medical personnel, interact with the data visualization tools we create?
- How can we make interactive tools more reproducible?
- How can we capture, store, and share user interactions? What can we learn from these interactions?
- How can we document research processes and decision-making processes?
- How can we guarantee that explanations in reports and publications are not post hoc rationalizing a biased decision? This applies to both human and AI decisions.
- How can we visually represent joint human and non-human (AI/ML) decision making?

What initially started off as a discussion on how we can capture and keep track of user interactions in interactive visualizations, quickly expanded to a discussion on how we can track, store and share the ways we arrive at insights in general. Many of us in the data

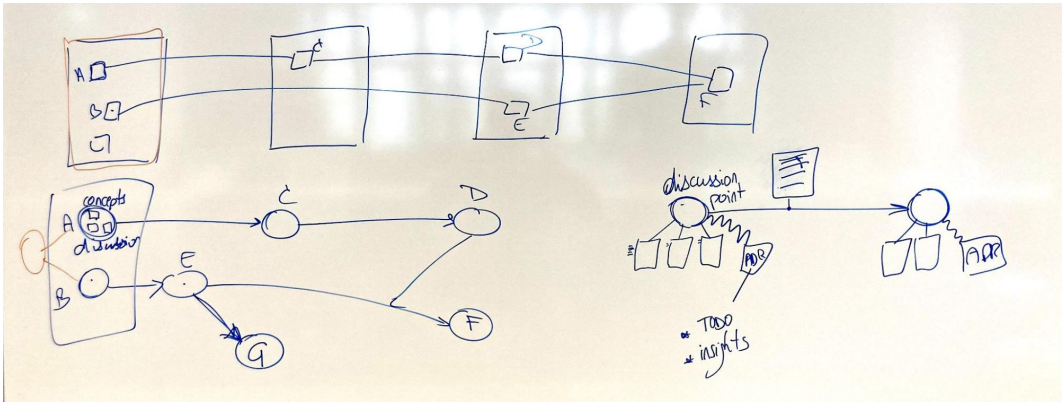

Figure 1 The provenance problem is portrayed as a network graph, in which the nodes represent decision points and the edges the processes that lead from one decision to the next (e.g. an algorithm or interactive notebook).

visualization field are accustomed to code-based notebooks, which are interactive, *in silico* equivalent of the biologist's lab notebook. We envisioned a tool that combines the best of both worlds: the interactivity, reproducibility, and scalability of the former and the flexibility and information density of the latter. We called this tool "the lab notebook of the future" (or, to lend it a certain futuristic air and for the sake of having a somewhat manageable deadline: "the lab notebook of 2030").

From a purely computational perspective, having provenance means that we need to record and store every step of the data visualization process. Any successful notebook must capture a sequence of views, insights, data points, and interactions that all together lead to a particular decision. There exist several technical approaches to this challenge, but we agreed that the flexibility of graph-based solutions makes them prime candidates (Figure 1). From a human perspective, having provenance means that we want to describe, explain, justify, adjust, and share the different steps taken to come to a particular visualization-driven insight.

High-quality provenance data allows for better retrospective exploration and analysis of provenance, which helps rescue the original intent of particular actions. It gives us a way to mine the cognitive trail created when using interactive tools. It also leaves an audit trail, which has several benefits in and of itself, from improving scientific reproducibility to boosting confidence in clinical decision making. Above all, provenance can help keep the agency in the visualization process. It ensures that a solution or visualization is arrived at actively, with intent. Interactive visualizations of (biological) data are inherently dependent on user input, as the visualization tool is not just delivering the same view regardless of user intent and actions.

We identified several hurdles that stand between us and perfect provenance. How do we, for example, deal with continuous, iterative analyses? Or layered hypotheses? Does provenance scale with the hierarchy and complexity of the data and knowledge, as in the case of biological data? Having provenance does not necessarily mean understanding it. How can we ensure that stakeholders have the same mental model when thinking about or discussing provenance? Depending on the complexity of an application, provenance data can quickly become unwieldy. They can also be difficult or impossible to interpret following system changes (e.g., software updates).

The end goal of introducing provenance in interactive data visualization tools is to empower the user. Achieving this goal will require a combination of human, algorithmic, and visualization solutions. The end goal of our Dagstuhl working group is to crystallize our discussions and ambitions into a (position) paper that describes the problem and details our proposed solution and to put them into practice by building "the lab notebook of 2030".

4.2 Recommendations for the design of visual, interpretable, and deep learning-based analytics pipelines in medical imaging

Katja Bühler (VRVis – Wien, AT), Guadalupe Canahuate (University of Iowa, US), Carsten Görg (University of Colorado – Aurora, US), Helena Jambor (Universitätsklinikum TU Dresden, DE), Georgeta Elisabeta Marai (University of Illinois Chicago, US), Timo Ropinski (Universität Ulm, DE), and Hagit Shatkay (University of Delaware – Newark, US)

Recent years have seen a constant increase in utility and multitude of medical images in clinical practice. Modern deep learning techniques have been developed and studied extensively in this context to cope with the availability and multitude of medical images. They promise to accelerate their use in clinical routines, like the diagnosis process and intervention planning and increase the reproducibility of clinical decision-making.

While the accuracy of these techniques is unmet, their applicability also leads to novel challenges. One such challenge is the interpretability of the AI decision process. When medical doctors are confronted with decisions made by AI, they are often not able to interpret the reasons behind such decisions. While various explainable AI techniques have been proposed to address this challenge, it is largely unclear how these techniques can be optimally used in the medical decision process. Deep learning experts design artificial neural network-based solutions that only focus on technical aspects, missing the end users' need for interpretability.

Visualization is a powerful communication channel for explainable AI, but existing visualization research is mainly decoupled from existing technical constraints and knowledge on what information can be delivered by XAI. Therefore, within this paper, we aim to build a bridge between both domains, providing a framework for visual XAI pipelines for image-based diagnostic workflows. We will initially review domain terminology as medical doctors communicate findings from medical images. Based on this terminology, we will then lay out so-called micro-tasks, as they are often solved by modern AI technologies, and link them to the needs of the medical experts. Finally, we will review recent explainable AI techniques and show how they can address these needs by visually communicating the results of the relevant microtasks.

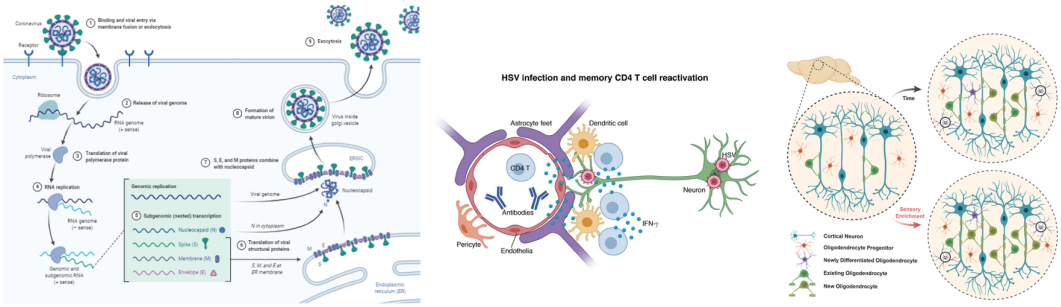

Figure 2 Examples of biomedical cartoons from `www.biorender.com`.

CARTOOMICS

Figure 3 Potential logo for the cartooning tool.

4.3 Semantically enabled biomedical cartooming

Lawrence Hunter (University of Colorado – Aurora, US), Emma Beauxis-Aussalet (VU University Amsterdam, NL), Nadezhda T. Doncheva (University of Copenhagen, DK), Lynda Hardman (CWI – Amsterdam, NL), and Martin Krzywinski (BC Cancer Research Centre – Vancouver, CA)

Originally constituted as a group to focus on semantic abstraction in bioviz, we split off from the main group to focus on a more specific task: Semantic support for drawing biomedical cartoons. Such cartoons are very common in molecular biology research publications and are used to summarize and communicate complex mechanistic hypotheses (Figure 2).

Existing drawing tools for producing such drawings and related interactive visualizations (e.g. Cytoscape) operate primarily on generic structures (e.g., networks, matrices) and provide domain-specific glyphs or templates (e.g., SBML or BioRender). Drawing operations are also generic (e.g., grouping or aligning drawing objects). They do not provide specific support to relate biomedical elements or concepts to the visual elements of the cartoons. Our idea is to exploit existing biomedical knowledge graphs to provide semantic support for a cartoon composition tool that would automatically link biomedical elements to templates of visual elements. We named such a tool Cartoomics (Figure 3).

Ontologically grounded biomedical knowledge graphs (e.g. PheKnowLator or SPOKE) open the potential for drawing tools to use existing knowledge graphs to help users create effective visualizations. We can use such visualization, for example, to explore the relationship of experimental results, compare it to prior knowledge, and communicate new findings. Starting from a set of experimental findings, the Cartoomics tools could exploit biomedical knowledge graphs to suggest visual elements to add to a cartoon ("semantic autocomplete"). Cartoomics could also allow the specification of semantically related entities to be manipulated as a group ("semantic templates", e.g., for aggregating or highlighting associated elements).

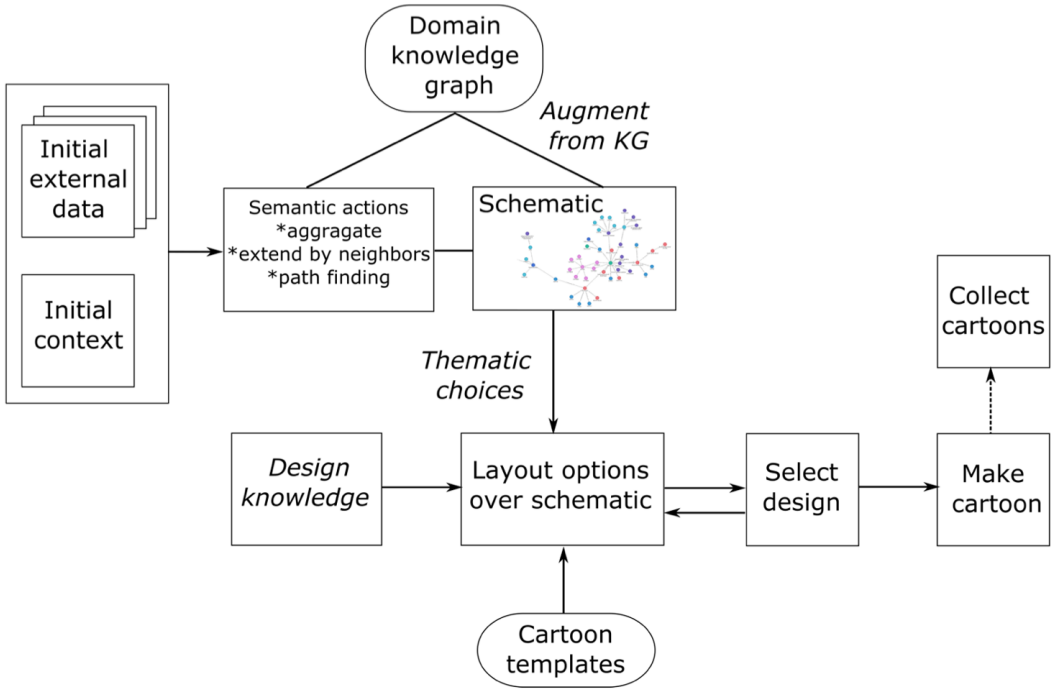

Figure 4 A conceptual diagram of a semantically enabled biomedical cartooning system.

Subsumption hierarchies in the ontologies that ground the knowledge graphs offer semantic zooming, e.g., for increasing or decreasing the level of detail. Other potential semantic actions include adding related concepts ("semantic neighbors"), following paths inferred from the knowledge graphs, embedding related concepts in aggregates, or exploring subgraphs with interactive queries.

We envisioned at least two possible application domains. In one, oncologists use genomic sequence data to make treatment decisions. Here, genomic data is combined with patient data and mapped to pathway diagrams, which are used to compare the likely effect of drugs. These diagrams vary from patient to patient but follow general patterns. The construction of these diagrams is time-consuming and requires expertise. Semantically enabled biomedical cartooning could potentially speed up the production of these diagrams and also improve their quality.

A second application domain would be interpreting transcriptomic, proteomic, or multi-omic experimental results. Here, a list of differentially expressed genes (e.g., proteins)needs to be understood (e.g., why these genes? how they relate to each other?). The investigation may combine contexts such as the description of an experimental manipulation (e.g. the addition of a compound). These elements need to be visually represented to communicate the main findings or hypotheses.

We drafted a conceptual design for such a system (Figure 4). The process begins from an existing domain knowledge graph and an initial input set of data and context (e.g., new experimental results). The tool would query the knowledge graph based on the initial input to produce a semantic schema, representing parts of interest in the existing knowledge graphs (e.g., with ontologically defined nodes and edges relevant to the inputs). The schema would be interactive and provide a variety of semantic affordances to modify the graph selected initially. Such modifications could include aggregation, extension by n-neighbors, pathfinding, subgraph matching, and even vector space aggregating neighbors using an embedding calculated from the graph.

When the user was happy with the semantic content of the selected schema and knowledge graphs, a second stage would solicit information about layouts of interest for the cartoons. We could draw the layouts from predefined cartoon templates or additional design knowledge graphs to recommend specific templates. The Cartoomics tool would allow the user to make thematic choices and to iterate through potential design alternatives. From there, automated processes could draw the final cartoons.

There was considerable enthusiasm among the participants for this idea. We plan to apply for grants, both in US and EU, to pursue the design and evaluation of the potential of such a cartooning system.

4.4 Visual Analytics of Multilayer Networks Representing Knowledge Graphs

Karsten Klein (Universität Konstanz, DE), Michael Behrisch (Utrecht University, NL), Andreas Kerren (Linköping University, SE), Stephen G. Kobourov (University of Arizona – Tucson, US), Michael Krone (Universität Tübingen, DE), Bruno Pinaud (University of Bordeaux, FR), and Falk Schreiber (Universität Konstanz, DE)

This working group originated from the discussions of the first breakout group on "data abstraction," which centered around abstraction for biomedical knowledge graphs. A *biomedical knowledge graph* is based on information about biological concepts and processes that are used to describe a complex biological system. It models corresponding biological relations between the involved entities and their effects on the system, such as protein interaction, functional associations between diseases and cellular processes, and gene expression. Due to the vast amount of information from life-science databases and publications, such a knowledge graph is usually huge, with millions of entities and relationships between them, resulting in the need for automated analysis and interactive exploration methods for knowledge extraction. Knowledge graphs can support a variety of biomedical applications, e.g., pharmaceutical or multi-omics applications [1].

The initial discussions of the "data abstraction" group centered around two different topics: (1) is a standard graph an appropriate abstraction for the underlying data, and (2) how to create the cartoon-like graphics common in life-science publications. Our subgroup first discussed general and conceptual aspects of data abstraction in the context of knowledge graph analysis to arrive at a common understanding. We defined a list of potential research projects that the group members are interested in, extending the original scope of data abstraction towards visual analytics of multilayer networks for the visual representation of knowledge graphs. The group consists primarily of visualization and network visualization researchers; hence our discussions were about appropriate abstractions for knowledge graphs, representation and visualization metaphors, network layouts, and algorithmic aspects. Notably, we agreed that multilayer networks (which capture different types of nodes and various types of relationships [2, 3]) might offer a more appropriate abstraction than standard graphs (made of one set of nodes and one set of links).

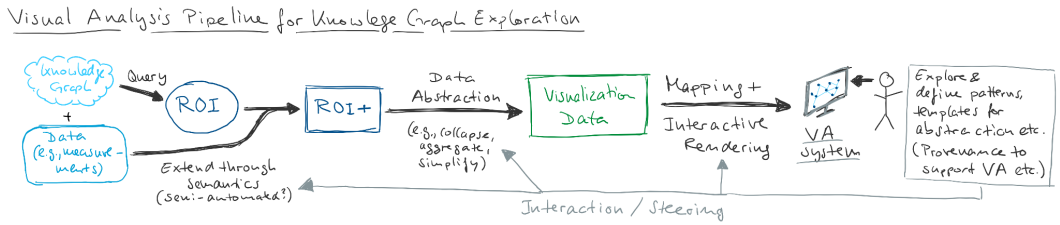

Figure 5 The visual analysis pipeline model used in the group's discussions.

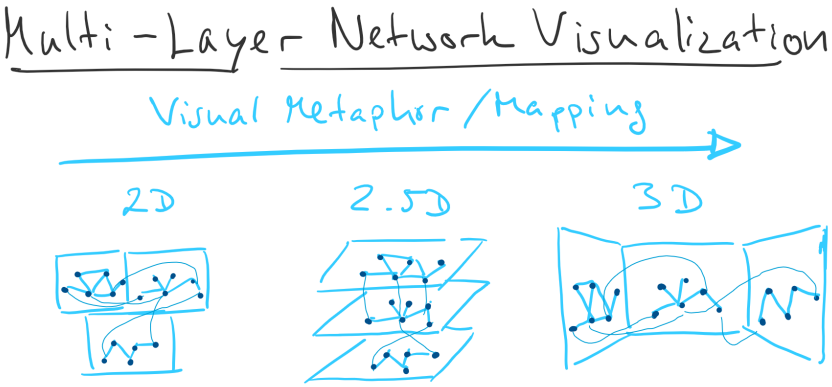

Figure 6 Visual metaphor classes for multilayer network visualization, comprising 2D, 2.5D, and 3D.

From the presentations of the domain experts, we derived a workflow model for knowledge graph analysis, which covers query-based subgraph extraction and semantic extension, abstraction, mapping on visualizations, and interactive exploration, see Figure 5 for getting an overview of this model. We defined several concrete research challenges that we briefly describe in the following paragraphs based on this workflow.

Challenge 1 – Visual Metaphors. This challenge investigates which visual metaphors are appropriate for representing multilayer networks, see Figure 6. Important characteristics of the metaphor include the dimensionality of the representation of the complete network (in 1D, 2D, 2.5D, or even 3D), the arrangement of layers, the encoding of entities, and abstractions. We want to design a human subjects study to evaluate different metaphors in combination with analysis tasks, which we can obtain from the existing task taxonomy described by McGee et al. [2, Chapter 4].

Challenge 2 – Interactive Exploration. The second challenge covers the impact of interaction and representation devices, e.g., immersive environments including 2D and stereo display, head-mounted display, and augmented/virtual reality [4]. Based on the results of our study for the first research challenge, we aim to compare different environment designs for one or more metaphors.

Challenge 3 – Network Analysis. The aspects of network analysis, which encompass centralities, motifs, aggregation, and clustering, and the need to incorporate the semantics, define our third challenge. Results here could inform the network visualization characteristics, including interactive layer arrangements, derivation, and filtering. Future work could include realizing a visual analytics system for knowledge graph analysis.

References

1 David N. Nicholson and Casey S. Greene. "Constructing knowledge graphs and their biomedical applications." Computational and structural biotechnology journal, 18, pp.1414-1428, 2020.
2 Fintan McGee, et al. *Visual Analysis of Multilayer Networks. Synthesis Lectures on Visualization.* 8(1), pp.1-150, 2021.
3 Falk Schreiber, Andreas Kerren, Katy Börner, Hans Hagen, and Dirk Zeckzer. *Heterogeneous Networks on Multiple Levels.* Multivariate Network Visualization: Dagstuhl Seminar #13201, Dagstuhl Castle, Germany, 2013.
4 Tim Dwyer, et al. *Immersive Analytics: An Introduction.* In Immersive Analytics, pp.1-23. Springer, Cham, 2018.

4.5 A curriculum for the future of bio+med-Vis

Torsten Möller (Universität Wien, AT), Emma Beauxis-Aussalet (VU University Amsterdam, NL), Helena Jambor (Universitätsklinikum TU Dresden, DE), Barbora Kozlíková (Masaryk University – Brno, CZ), Kay Katja Nieselt (Universität Tübingen, DE), and William Ray (Ohio State University – Columbus, US)

The communication role of visualization is well accepted in the bio+med communities. At the same time, it is essential to make biological and medical practitioners more aware of the value of visualization, not only as a dissemination tool but also as a data analysis tool. The development of a Curriculum for these communities, bio+med, and visualization, has thus a high priority and aims to familiarise students with modern visual analysis methodologies applied to biological and medical data and provide hands-on training.

While several visualization community members are teaching summer camps, tutorials, and workshops on biological and medical data visualization, many of these educational sessions take the form of an introduction to specific tools. We find ourselves handling similar questions: What is exploratory data visualization? What is visual analysis, which frameworks to think about visualization exist, how can we explore design space? And how can we visualise biological and medical data to gain insights into them so that hypotheses can be generated, explored, and validated, and further analyses can be targeted for future work?

Despite the increasing importance of visualization for the bio+med communities, there is currently a general lack of integration into curricula, and hence a lack of visual literacy. A useful and appropriate curriculum has not yet been developed. Therefore, in our group, we focused on addressing the following questions:

- What should a modern and seminal curriculum for visualization for the bio+med communities contain?
- With the different target audiences in mind, how far along in introducing visualization should such curriculum go while also integrating courses on the domain and topics on data?
- What are the essential topics for such a curriculum, and how can it provide comprehensive hands-on training?

The actions taken in this working group were shaped based on these questions and on the previous initiative of building a teaching platform for the bio+med+vis audience.

The latter developed into the first Bio+Med+Vis Spring School, held in 2021 (`https://biomedvis.github.io/`). The initial ideas for this initiative were drawn from the 2018 BioVis Dagstuhl seminar 18161 "Visualization of Biological Data – Crossroads", and from the 2020 Shonan seminar 167 on "Formalizing Biological and Medical Visualization".

Within our working group, we aimed to reflect on and review what has been done so far to extend and enhance the matrix of topics for the curriculum. The current status of the matrix can be found following this link.

We first revised the current content in designing the curriculum and discussed potentially missing areas or topics. We then extended the overall *Platform* table by the *Overview* row, which will serve for collecting introductory topics and prerequisites for individual methods (listed in table columns), without any specifics of scales (listed in table rows). We further added the *Visual Design* table with its corresponding content. We also added the *Contributors* section, where we asked all participants of the Dagstuhl seminar to express their willingness to participate in this initiative. Our goal was to identify enthusiastic collaborators and potential driving forces of the upcoming initiatives in this context. Our working group also identified the need for defining the learning outcomes for each module, as these are one of the driving forces for the content of the curriculum.

The next actions that will follow after the Dagstuhl seminar will be:

- Finalizing the matrix content by filling in the missing gaps in individual tables.
- Addressing experts in given fields to prepare the educational materials for topics listed in the matrix.
- Initiating the creation of the BioMedVis educational platform to collect the educational materials and make them publicly available.
- Addressing potential collaborators regarding the preparation of the following Spring School and discussing the date and content with them.
- Explicitly joining forces within the BioVis + VCBM organizational structures to bring the communities together and strengthen our research and educational efforts.

Finally, we would like to acknowledge the collaboration of our colleagues not attending the seminar but who have contributed to the results of our working group. These include Jan Byška, Renata Raidou, Noeska Natasja Smit, Ingrid Hotz, Johanna Beyer, and Hsiang-Yun Wu.

4.6 Facilitating cross-expertise exploration in XAI for multi-omics via visualization

Cagatay Turkay (University of Warwick – Coventry, GB), Jillian Aurisano (University of Cincinnati, US), Theresa Anisja Harbig (Universität Tübingen, DE), Raghu Machiraju (Ohio State University – Columbus, US), and Mathias Witte Paz (Universität Tübingen, DE)

With the emerging high-throughput methods in biology, we can systematically collect the data from multiple omics levels such as genomics, transcriptomics, proteomics, and metabolomics. Therefore, in addition to analyzing single entities, such as a small group of genes in one omics layer, researchers can now analyze larger sets of entities across multiple omics layers to increase their understanding of complex biological systems. However, the produced data

of all layers need to be analyzed simultaneously for cross-features such as (anti-)correlations or common clusters. Hence, there is an increasing interest in applying artificial intelligence (AI) methods to not just one but several combined "omics" datasets. Since data experts are not often familiar with AI methods, a further actor comes into play for method development. With this, a new challenge within the omics field arises since the domain experts must communicate their research approaches with non-experts, and model developers also need to share their knowledge. However, the combination of the data and the applied methods complicates the explainability of the results since both are already complex by themselves. We believe that we can use visualization techniques to facilitate the communication between actors.

Here, we first discuss the characteristics of multi-omics data that make the development of AI methods more difficult. We then identify the current stakeholders within this AI-multi-omics system and propose analyzing the existing visualization techniques via a survey.

Multi-Omics Data Characteristics

Multi-omics data have particular characteristics that can complicate AI model development and make it difficult to explain its results. First of all, multi-omics data is high-dimensional. Depending on the study, genomics, transcriptomics, proteomics, metabolomics, and/or epigenomics data can be included (among others). Analyzing and understanding these data types on their own has been a longstanding challenge for the field since high throughput measurements were introduced to molecular biology research, let alone understanding the combination of multiple omics data, or the results of an abstract AI model, or the AI model itself. AI methods have tremendous potential to enable researchers to make sense of complex data. However, the problem of helping domain experts understand these systems is complicated by the use of opaque AI methods.

Several challenges in multi-omics data creation and collection complicate model development and interpretation. First of all, multi-omics data sets can be very different. The number of samples and the analyzed omics levels can vary immensely, leading to different requirements for the model depending on the data. Moreover, missing data is a challenge for all omics data. For example, in genomics, not all genes might be annotated, or the functions of some genes might be missing, which can lead to an incomplete model or difficulties in understanding the model results. Moreover, missing data can lead to data imbalance. For example, when analyzing cancer data, certain cancer types are much more common than others and characterized more thoroughly. Therefore, when interpreting the results of an AI model or the model itself, it is always important to be aware of the quality of the input data. If the results do not match the expectations when using this input data, it does not necessarily mean that the model is flawed or that the expected signal could not be detected. It could simply mean that the input data was not sufficient.

Finally, several challenges arise when a researcher interprets the result of an AI model or the model itself. Multi-omics is an emerging field, and domain experts often specialize in one specific omics data type. Naturally, this can lead to an interpretation of the data and the artificial intelligence results' centered on a particular field of omics. Domain experts might attempt to apply AI techniques to increase the understanding of an already known phenomenon or to find an explanation for a hypothesis about the data. In contrast, non-domain experts might have a completely different, more unbiased view of the data and perform a purely data-driven analysis more often. However, they tend to have less background expertise and less knowledge about potentially missing data. This can lead to a challenge in communication between different stakeholders dealing with multi-omics data.

Stakeholders

From our experience in the multi-omics field, we have distinguished stakeholders with different goals and interests concerning the AI methods applied for the data analysis. Commonly known stakeholders from this field are the modelers and the domain experts. Even though this division might not always be as sharp as described here, it is common for these two actors to have disjoint backgrounds. Usually, the modelers tend to have less domain expertise from the applied field (e.g. microbiology). In contrast, domain experts might have little to no understanding of the underlying machine-learning models. This could lead to the different interests of both parties: the modelers might be primarily involved with the model's performance rather than with the underlying phenomena that the model captures. At the same time, the domain experts focus only on the model's final outputs since understanding the decision path would require time and background expertise they usually do not possess.

However, there might be situations where these two stakeholders require detailed knowledge on the complementary side of the system, especially when a problem with the model occurs, such as a difference in the expected output from the side of the domain expert or low performance of the model after being evaluated by the modeler. For this, they usually go to an explainer or assume such a role for having an accurate translation between model development, data expertise, and artificial intelligence methods. By integrating both sides of the project and encoding the missing knowledge in a simplified way, such as in a visualization, the person assuming the role of the explainer could also motivate the two stakeholders to be curious and explore the decision paths of the model. We believe more research is needed to develop compelling visualizations that facilitate the task of the explainer acting as a translator and motivate the stakeholders to explore the decision paths by themselves in more detail, even if no issue has been detected in the developed model. Such curious stakeholders might be able to find information that could result in breakthroughs for their corresponding field by accounting for the model's output or structure and considering the model's behavior.

Outlook

Though the previously described stakeholders were identified from our prior experiences in the multi-omics field, it is unclear how often the actors appear, especially the explainer. To clarify this, we propose to develop a survey on how often curiosity comes into play in a multi-omics project applying AI methods, especially concerning the decision paths taken by the model. We also expect to gather current visualization techniques used from the explainers' side to reduce the gap between the parties. The gathered insights will help develop future tools that apply machine-learning methods in multi-omics data and avoid curious actors requiring the help of an explainer.

Nevertheless, it might be possible that the number of curious actors is not high enough for us to collect sufficient materials. This would mean that, although theoretically reducing the gap between modelers and domain experts would provide a better insight into the studied phenomena, such approaches may not be feasible or desirable in reality. In such cases, our focus from our survey results would be different: we would start thinking of visualization techniques that motivate the stakeholders to think outside their boxes and consider the whole picture. We believe that by reducing such a gap, the dogmas of both sides might be questioned more intensely, and it might facilitate the multi-omics data to provide new perspectives into the applied fields.

Participants

- Jan Aerts
Hasselt University – Diepenbeek, BE
- Emma Beauxis-Aussalet
VU University Amsterdam, NL
- Michael Behrisch
Utrecht University, NL
- Katja Bühler
VRVis – Wien, AT
- Nadezhda T. Doncheva
University of Copenhagen, DK
- Mennatallah El-Assady
ETH Zürich, CH
- Carsten Görg
University of Colorado – Aurora, US
- Theresa Anisja Harbig
Universität Tübingen, DE
- Lynda Hardman
CWI – Amsterdam, NL
- Lawrence Hunter
University of Colorado – Aurora, US
- Helena Jambor
Universitätsklinikum TU Dresden, DE
- Andreas Kerren
Linköping University, SE
- Karsten Klein
Universität Konstanz, DE
- Stephen G. Kobourov
University of Arizona – Tucson, US
- Alexander Koch
BioLizard – Gent, BE
- Michael Krone
Universität Tübingen, DE
- Martin Krzywinski
BC Cancer Research Centre – Vancouver, CA
- Alexander Lex
University of Utah – Salt Lake City, US
- Lennart Martens
Ghent University, BE
- Torsten Möller
Universität Wien, AT
- Kay Katja Nieselt
Universität Tübingen, DE
- Bruno Pinaud
University of Bordeaux, FR
- Timo Ropinski
Universität Ulm, DE
- Falk Schreiber
Universität Konstanz, DE
- Cagatay Turkay
University of Warwick – Coventry, GB
- Blaz Zupan
University of Ljubljana, SI

Remote Participants

- Jillian Aurisano
University of Cincinnati, US
- Tanya Berger-Wolf
Ohio State University – Columbus, US
- Katy Börner
Indiana University – Bloomington, US
- Andreas Bueckle
Indiana University – Bloomington, US
- Guadalupe Canahuate
University of Iowa, US
- Jian Chen
Ohio State University – Columbus, US
- Nils Gehlenborg
Harvard University – Boston, US
- Barbora Kozlíková
Masaryk University – Brno, CZ
- Raghu Machiraju
Ohio State University – Columbus, US

- Georgeta Elisabeta Marai
University of Illinois
Chicago, US
- James Procter
University of Dundee, GB
- William Ray
Ohio State University –
Columbus, US
- Jos B.T.M. Roerdink
University of Groningen, NL
- Ryo Sakai
GSK – Uxbridge, GB
- Hagit Shatkay
University of Delaware –
Newark, US
- Charlotte Soneson
FMI – Basel, CH
- Miha Štajdohar
Genialis – Houston, US
- Marc Streit
Johannes Kepler Universität
Linz, AT
- Granger Sutton
J. Craig Venter Institute –
Rockville, US
- Danielle Szafir
University of North Carolina at
Chapel Hill, US
- Mathias Witte Paz
Universität Tübingen, DE