



HAL
open science

Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches

Mashaël Alkadi, Vanessa Serrano, James Scott-Brown, Catherine Plaisant, Jean-Daniel Fekete, Uta Hinrichs, Benjamin Bach

► **To cite this version:**

Mashaël Alkadi, Vanessa Serrano, James Scott-Brown, Catherine Plaisant, Jean-Daniel Fekete, et al.. Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches. IEEE Transactions on Visualization and Computer Graphics, 2023, 29 (1), pp.907-917. 10.1109/TVCG.2022.3209487. hal-03738500

HAL Id: hal-03738500

<https://inria.hal.science/hal-03738500>

Submitted on 26 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches

Mashaël AlKadi, Vanessa Serrano, James Scott-Brown, Catherine Plaisant,
Jean-Daniel Fekete, Uta Hinrichs and Benjamin Bach

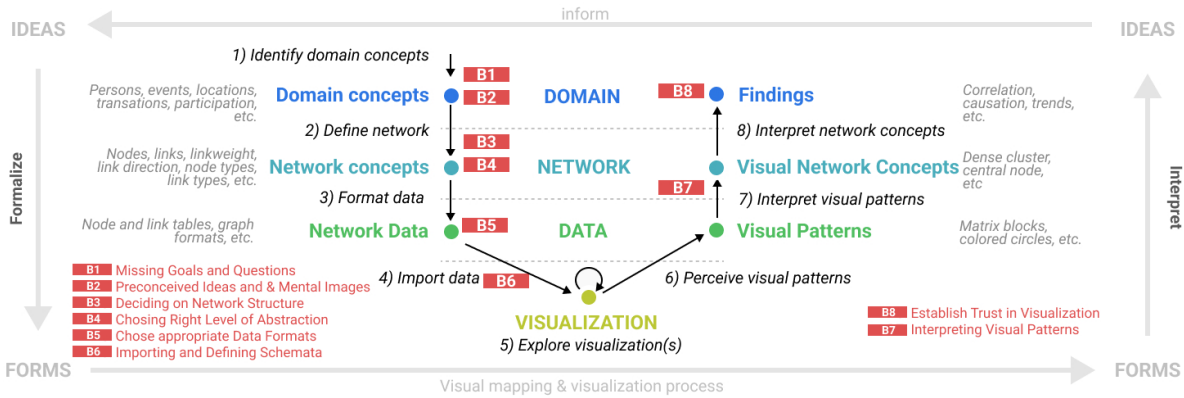


Fig. 1: Barriers (red) that may be encountered during visual network exploration process, while translating domain concepts (ideas) into network structures and visualizations (forms) and back into findings.

Abstract—This article reports on an in-depth study that investigates barriers to network exploration with visualizations. Network visualization tools are becoming increasingly popular but little is known about how analysts plan and engage in the visual exploration of network data—which exploration strategies they employ, and how they prepare their data, define questions, and decide on visual mappings. Our study involved a series of workshops, interaction logging, and observations from a 6-week network exploration course. Our findings shed light into the stages that define analysts’ approaches to network visualization and barriers experienced by some analysts during their network visualization processes. These barriers mainly appear before using a specific tool and include defining exploration goals, identifying relevant network structures and abstractions, or creating appropriate visual mappings for their network data. Our findings future work in visualization education and analyst-centered network visualization tool design.

1 INTRODUCTION

Network analysis is becoming an established methodology across many disciplines such as biology, the social sciences, and the humanities. While the specific datasets are diverse and heterogeneous with respect to how they have been captured and what research questions they aim to address, they have in common that the analyst thought of their data as a *network*—i.e., as nodes and links. By applying methods from network analysis, such as calculating metrics or producing a visualization, the analyst aims to learn something about these nodes and their relationships. Hence, network analysis is a *tool* and a *lens* to model and interrogate data. It comes with assumptions, such as that a network is a meaningful representation for the data. And, it involves multiple steps in defining nodes and links, formatting data, applying visualization, and interpreting these visualizations correctly.

Especially for novice analysts, performing these steps and making the respective decisions can pose problems. Over several years, we

- Mashaël AlKadi, James Scott-Brown, Uta Hinrichs, and Benjamin Bach are with University of Edinburgh. E-mail: {bbach}@ed.ac.uk.
- Vanessa Serrano is with Ramon Llull University
- Catherine Plaisant is with University of Maryland and Inria. E-mail: plaisant@umd.edu.
- Jean-Daniel Fekete is with Université Paris-Saclay, CNRS, Inria. E-mail: Jean-Daniel.Fekete@inria.fr.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxx

ran workshops, tutorials, talks, and interdisciplinary collaborations on network visualization and built our own network visualization tool, the Vistorian [3, 7]. During this time, we experienced first-hand some of the problems, challenges, and misconceptions that can hamper engagement with network analysis and visual network exploration. Many of the issues we encountered were very common, such as identifying the meaningful entities (nodes, links, attributes, time, geographical locations) from the initial dataset—e.g., a corpus of documents—that should represent nodes and links to create a network that meaningful to answering a specific research question. Other issues we observed were about using visualizations other than node-link diagrams to understand multivariate, dynamic, and geographic networks, or properly formatting a dataset for import into a tool.

While some of these issues are echoed in studies on expert data workers [8, 15, 42, 51] or in the context of network visualization literacy [18, 59, 65], a comprehensive account of the problems novices face when engaging with network exploration is missing. Whereas multivariate and quantitative data can be represented by familiar visualizations (e.g., barcharts or line charts), networks pose specific challenges due to their qualitative (i.e., structural) character, the combination with other data types, and the abundance of specialized visualization techniques for temporal [12], geographic [67], and multivariate [53] networks.

In this article, we present an in-depth study into the barriers some analysts face when preparing their data for visual exploration and using an interactive visual network exploration tool. For simplicity, we relied on a single tool, the Vistorian, that we consider a state-of-the-art network visualization tool comparable to Gephi [9], NodeXL [71, 72], or Palladio [73]. The Vistorian has been specifically developed to provide an easy entry point to network visualizations (node-link, adjacency matrix, timeline, map) and interactive exploration for data analysts

without a technical background (Sect. 3). In a first log study (Study 1, Sect. 4) we tracked users of the Vistorian over several months with the aim of understanding how they use visualizations and interactions in their day-to-day work. To complement these anonymous data with qualitative data about potential barriers, we designed a 6-week network exploration course and delivered this to 36 analysts whilst closely monitoring their progress, giving personal advice, and engaging in individual interviews (Study 2, Sect. 5). The course was open to anybody and included analysts from computer science, the social sciences, and business. It introduced the basic concepts of networks, example visualizations for multivariate, temporal, and geographical networks, hands-on activities for sketching and formatting data, and hands-on tutorial with the Vistorian.

While many of our participants succeeded in visual exploration and did not encounter major problems, our observations reveal barriers, experienced by some participants. For example, preconceived ideas and specific mental images about one’s network blocked analysts from thinking creatively about their research questions and network structures. We also found a barrier in defining (multiple) possible networks to match different research questions and abstract from a domain problem into a network problem and back [20,52]. Mitigating these barriers has implications for future tools and research. In summary, we report:

- the user profiles we identified based on anonymous interaction logs across 56 weeks (Sect. 4),
- observations (Sect. 7) from a 6-week course (Sect. 5) on visual network exploration, including
- a set of detailed participant profiles and observations about goals and strategies toward network visualization and exploration (Sect. 6),
- 8 barriers to engaging with network exploration (Sect. 9)
- implications for tool design and training (Sect. 10).

2 BACKGROUND

Tools for Visual Network Exploration—Numerous tools enable analysts to produce network visualizations without writing any code: Gephi [9], SoNIA [13], Palladio [73], NodeXL [71,72], Visone [11], NetworkWorkbench [56], SYF [58], Paejk [27], UCINET [64], or Net-Miner [26]. These tools provide node-link diagrams with some degree of interactivity, maps with overlaid node-link diagram; or animation support for temporal networks. These tools load data in a range of formats, such as GraphML, GEDCOM, GEFX, etc. [63], some of which are not human-readable or manually editable. NodeXL asks users to upload data as *node tables*, which contain information about nodes, and *link tables*, which describes links between these nodes. Users then specify the meaning of each column and perform transformations such as link projection. That said, data-wrangling tasks such as re-formatting data and correcting errors in ‘dirty’ datasets can be a significant portion of data-analysis work [41]. Dedicated tools can be used for data wrangling, for example, Ploceus [46], Orion [36] or Origraph [14].

Our own network visualization tool, the Vistorian, is representative of these tools in that it creates interactive node-link diagrams visualizing link weight, direction, and type as well as node types, and can represent temporal and geographical data. It had been developed and iterated over with many domain collaborators and allows us to log interaction data during in-the-wild and day-to-day use. We were inspired by NodeXL’s table format, which is easy to create and edit in common spreadsheet applications. At the same time, the Vistorian extends existing tools by providing alternative visualization techniques such as adjacency matrices (e.g., for dense networks), a timeline and timesliders for temporal networks, a map for geographic networks, and coordinated linked views.

Studying Network Visualizations and Tools—Many studies [78] have empirically evaluated the value of specific visual representations for networks (e.g. [31, 38, 54, 57, 62, 66]), their respective visual affordances (e.g., [1, 2, 61, 74]), interaction techniques (e.g., [6]) or the combination of multiple coordinated views for node-link and matrix diagrams [22,37]. Due to the controlled nature of these studies and questions they pursue, they have often relied on tasks described as *low-level* by common task taxonomies [43, 44, 77]. Most of these studies focus on quantitative records to describe task-completion time and task-error

rate, but alternative measures such as learnability, user satisfaction [60], and engagement (e.g., [39, 47]) can be highly important.

To achieve greater ecological validity and investigate tasks that are higher-level and potentially more sophisticated, studies have used more complex methodologies and data collection mechanisms, such as studying analysts using design-studies [68] or involving experts and their data in qualitative studies [58]. Still, many of these studies focus on developing, understanding, and improving a specific visualization tool within a particular context. The Multidimensional, In-depth, Long-term Case (MILC) approach [70] streamlines the process of in-depth evaluation of visualization tools in the wild, and suggests using interviews, training, improving the tool, interaction logging [28] and ethnographic observations, collecting user annotations through logbooks and micro-entries to help tracking analysts thoughts and discoveries [16], as well as entry and exit questionnaires. In our pilot study (Sect. 4), we collected user logs over several months, supported by short mini-questionnaires as described by Serrano et al. [69]. The mini-questionnaires prompted users to provide feedback on their use of our tool to complement the interaction logs [33]. In our second study, we follow more closely the MILC methodology, but significantly increase participant numbers compared to prior studies (usually 5 [70]) to approx. 24 in our case.

Network Visualization Literacy and Training—Through brainstorming with scientists, teachers, and students, Sayama *et al.* [65] define seven “*essential concepts about networks [...] that every person in the 21 century [should] know by the time he/she finishes secondary education*”. The concepts include “*networks describe how things connect and interact*”, “*networks help reveal patterns*”, and “*visualizations can help provide an understanding of networks*”. However high-level these concepts seem, studies on teaching network science in high schools and public exhibitions [18, 25, 79] revealed how little is generally known about the basic concepts of networks, their analysis and visualization in the wider population. For example, when shown images of node-link diagrams, visitors to science museums often struggled to explain how these images would be interpreted [18].

Courses and explicit training for network analysis [25] have been used in the context of MILCs to evaluate NodeXL and learn about teaching network exploration concepts [17, 35]. These studies introduced graduate students (21 and 15 respectively) to Social Network Analysis and NodeXL and tracked them over 5 weeks. Students used demo datasets and smaller self-chosen datasets, described as *discovery learning* and *exploratory learning* in the paper. Our course (Sect. 5) focused on practitioners working on their own datasets, a stark qualitative difference to prior courses and studies. Moreover, rather than evaluating our tool (the Vistorian), we want to better understand the challenges that analysts during the network exploration process.

3 NETWORK VISUALIZATION WITH THE VISTORIAN

We describe the features of the Vistorian relevant to this research below.

Visualizations and Interactions—The Vistorian features four interactive visualizations (Fig. 2) to show different facts of a network if present. Each visualization provides pan and zoom capability and a legend showing node and link types, which can be used to filter nodes and links. If the network has a temporal component, a time range-slider is shown and allows the view to be filtered to a specified period. Individually, each visualization has a set of specific visual encodings and parameters that can be set by a user. For example, the **node-link diagram** features small sliders to set node size and opacity, link width and opacity, and the distance between links (lines) connecting the same two nodes (a value of 0 visually overlays all those links, collapsing them into a single line). The **adjacency matrix** can be ordered alphabetically or using a reordering method [30]. A cell can encode link type (color), link weight (opacity), multiple links (splitting the cell into vertical segments), and link direction (a gradient of increasing color points in the direction of whichever of the row or column represents the target node). The **timeline visualization** shows nodes along a vertical axis and links between nodes as arcs, along a horizontal time axis (Fig. 2). A user can zoom into a time period and clicking a node, keeps only this node’s neighbors visible, effectively compressing the view. The **map visualization** overlays a node-link diagram onto a

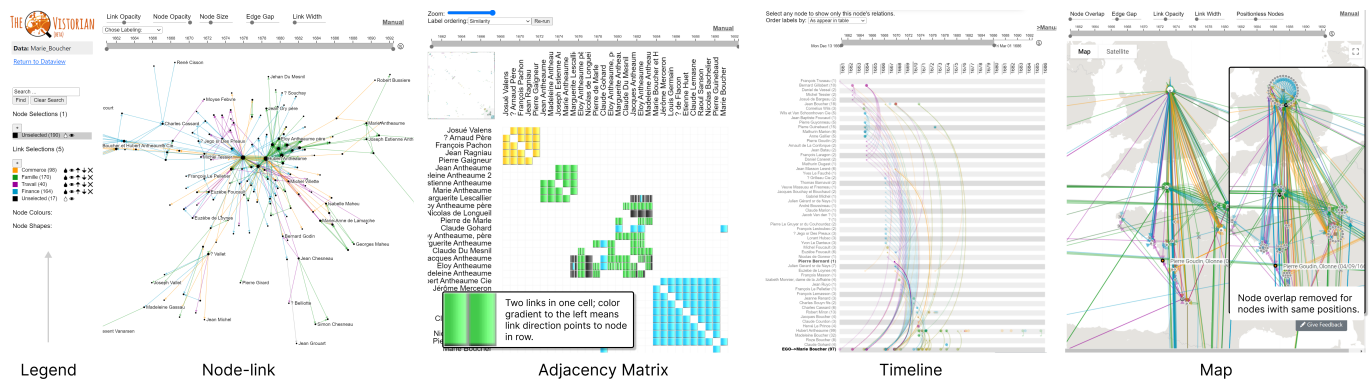


Fig. 2: **Visualizations in the Vistorian:** adjacency matrix, node-link diagram, timeline, and map with overlaid links and edges. Each visualization runs in its own tab coordinated through brushing and linking.

geographic map (Google Map API). Nodes with the same geographic position can be shown in a circle layout to make all their links visible. The circle radius can be reduced to collapse of all nodes into a single visual node, or increased to show all nodes. Each visualization in the Vistorian runs in its own browser tab, allowing tabs to be arranged in a coordinated-view or across multiple screens to access more than one visualization at the same time. Interactions, such as highlighting with brushing-and-linking, filtering time through a time-slider, and filtering the visibility of node and link types, are synchronized across tabs and visualizations.

Data Import—The Vistorian’s users upload data as node and link tables, inspired by NodeXL. Each row of a *link table* defines a link, with required columns for the source and target node, and *optional* columns for additional information on link weight, link date/time, link type, and geographic locations of the source and target node. Each row of a *node table* defines a node, and columns can specify the display name and node type. Locations in the link table can be expressed as place names (e.g., Paris) and are geocoded using the MapTiler API [48]. Upon upload, users can specify the network schema by defining a column’s interpretation from a drop-down list for each column (Source Node, Target Node, Link Weight etc.) Again, this feature was inspired by NodeXL. As soon as all required schema fields (only source and target node) for a network are specified, a status message above the table informs the user that the network is ready for visualization. Our website vistorian.net contains demo videos showing data upload and exploration, a detailed explanation on formatting data in node and link tables, links to fully-functional visualizations with a demo-dataset pre-loaded, and a FAQ.

Studying Real-World Usage of the Vistorian—While designed for analysts in the social sciences and humanities, it addresses a potentially wide audience. It has been designed and optimized with collaborators in these areas since 2014. It supports exploration processes that can easily take weeks, months, or years in which an analyst iteratively obtains new data, refines their research questions and explores their dataset from different perspectives. In such scenarios, an analyst’s understanding of the dataset is constantly evolving alongside their analysis goals, questions, and methods. Data exploration is almost always a process that requires learning and mastering new methodological tools, including visual exploration. Understanding these processes and how to facilitate them is the main aim of our research. Our studies of real-world usage of the Vistorian consists of two phases that build upon each other, as detailed in the following two sections.

4 STUDY 1: INTERACTION LOGGING & WORKSHOPS

The first phase of our research focused on anonymously logging interactions and conducting workshops to gain a general understanding of *how real-world analysts make use of the Vistorian and how to introduce analysts of various backgrounds to network visualization.*

Interaction Logging— we anonymously tracked all interactions with the Vistorian. Over 10 months, we collected data on upload and schema specification, creating and interacting with a visualization, and requesting help. A pop-up informed people about the logging taking

place, linked to information for study participants, and obtained their consent. This procedure was approved by our University Ethics Board. Logging started from the moment a user landed on the Vistorian website and continued until they closed all related browser pages. Logs were collected using the *Intertrace* JS library [29]. To facilitate the analysis of our interaction logs, we classified all log events, based on categories we defined in advance. The main categories (the list of all categories is included in the supplemental material) included *Data-Management* events (e.g., interactions such as creating a new network, uploading data tables, updating table entries, specifying table schema), *Visualization & Interaction* events (e.g., any interactions with visualizations), *Support & Communication* events (e.g., engagement with the Vistorian help resources such as manuals, demo data and videos), and *Bookmark* events (e.g., interactions related to the creation, update, and usage of bookmarks and annotations).

Workshops—While still developing the Vistorian platform, we ran a number of public workshops that varied in length from half a day to three days. Notable here are the most recent (online) workshops that we ran in parallel to starting to log interactions with the platform—one ran in the context of an online DataVis Meetup (approx. 25 participants) and one at an online Digital Humanities Summer School (approx. 55 participants). These workshops included lectures on multivariate network visualization, sketching networks and discussing appropriate visualizations, demos of the Vistorian, and hands-on tutorials that guided participants through the process of importing their own or a demo data to the platform. Data exploration using the Vistorian was facilitated at the workshops, but also beyond through drop-in sessions and a dedicated Slack channel where people could ask questions and seek support for their data explorations. While these workshops did not include structured data collection, they provided valuable insights that contextualized our analysis of interaction logs, and informed the network visualization course design (Sect. 5).

Findings from Interaction Logs & Workshops— Over the course of 10 months we logged 534 individual users sessions which we analyzed using the interaction categories described above, looking for interaction patterns (i.e., interaction times, what parts of the Vistorian platform people engaged with and in what order, and the number of returning users). Based on our analysis, we identified four types of usage patterns: **Demo-users** (298/534 \approx 55%) solely played with the demo dataset or looked into help resources, without exploring their own data using the Vistorian platform. **Data-strugglers** (71/534 \approx 13%) did try to explore their own datasets using the Vistorian platform, but struggled with shaping their data and creating their networks. These users sometimes returned to the platform several times, but without managing to successfully create a network visualization. **Single-session explorers** (93/534 \approx 17%) managed to upload their own data and created a network visualization. They then spent a short time (less than 2 min.) exploring their visualization and did not return. **Multi-session explorers** (72/534 \approx 13%) uploaded their own data, created (one or more) network visualizations and explored these over the course of multiple sessions that each lasted a minimum of 2 minutes. Multiple short sessions were more common than long continuous sessions.

Our results confirm observations we made during earlier workshops and demos. However, the mini-questionnaires did not help generate detailed insights about users' strategies and their problems in using the tool. We wanted to learn more about people's network visualization and exploration strategies and inherent issues that seem to go beyond potential usability problems as many people managed to upload and visualize their data using the Vistorian. To address these questions from a more qualitative data perspective, we designed the course for Study 2.

5 STUDY 2: NETWORK EXPLORATION COURSE

The target audience of our network visualization course were analysts without a technical background and/or no or little experience in network visualization and exploration. We advertised the workshop widely, targeting analysts across the sciences, arts and humanities, and at different stages of their analysis process. We aimed at teaching different aspects of network exploration—from data preparation, to visual mapping, to visual network exploration. We utilized the Vistorian as the sole network visualization and exploration tool in our course (1) to avoid overwhelming participants by introducing them to multiple network visualization tools within a relatively short amount of time, and (2) because the Vistorian supports a large variety of network visualization and exploration techniques while providing a relatively low entry point related tasks, especially for analysts without a technical background. Based on our findings from Study 1, we improved some features of the Vistorian. Below we describe these changes to the Vistorian, our course design and our methods of data collection and analysis.

5.1 Data Import Wizard

We replaced the 'manual' process of setting up a network schema with a comprehensive wizard, to support analysts importing the correctly formatted data and defining meaningful schemata. The wizard (see supplementary material for a figure) has the following steps: (1) network name, (2) data format (tables, GEDCOM, Pajek, etc.), (3) choosing between a link or a node table, (4) uploading a table and specify a schema for that table, (5) specifying a location and/or a node table with node types (optional), and (6) launching a visualization in a new tab.

5.2 Course Outline

We designed a 6-week network visualization course that covered both data preparation and visual network exploration. From a teaching perspective, the goals of the course were to help participants (1) Define the goals of their data exploration, (2) Become familiar with a range of network visualization techniques, through theory and hands-on use, and (3) Use different types of interactive visualizations to explore their own data. From a research perspective, we aimed to learn more about analyst approaches and strategies when using network visualization to explore their data and potential barriers inherent in these processes.

Following the recommendation by Batch *et al.* [10] to increase the use of visualizations in data exploration by providing education, the course was based on problem-based learning. We encouraged participants to bring their own data and related questions/issues/problems to the course, so they could directly apply lessons-learned in a real-world context relevant to them. However, we also provided demo datasets to those who did not (yet) have their own data.

Each week consisted of a two-hour session that combined online live lectures (recorded and shared online) with hands-on activities. Some sessions set a homework for the following week, to encourage participants to practice and identify the areas that they needed help and support with. We offered weekly one-hour drop-in sessions during which course participants could ask questions about the lecture material or raise issues they had encountered. We also moderated a Slack channel that allowed participants to discuss course-related questions and problems with each other and with the course instructors. We describe the topics and activities covered in the individual course sessions below.

Weeks 1 & 2: Network Data Preparation introduced participants to the topic of network visualization and the goals of the course. We covered the basic terminology around network visualization (e.g., introducing the concept of a network, nodes and links) as well as questions that can be answered using network visualization and exploration. In

Activity 1, participants prepared a sketch of the potential network(s) that could be constructed from their data and to think about potential questions that these networks could help answer. In **Activity 2**, participants identified attributes in their data that could serve as nodes and links in their data. In **Activity 3**, participants sketched small networks based on the nodes and links identified in Activity 2. In **Activity 4**, participants formalized these sketches into a concept map, by starting to define nodes and links more formally. Finally, in **Activity 5**, participants specified their node and/or link tables. We had initially planned for a single week to network data preparation, but during Week 1 we realized that participants struggled with (1) mapping data attributes to nodes and links to structure their network and with (2) formatting their data into a network table. During Week 2, we therefore dedicated more time to re-iterating on and finishing the activities introduced in Week 1.

Week 3: Data Shaping Techniques & Challenges introduced techniques for ensuring consistency in data, reducing data size (e.g., filtering nodes by degree or node type, or filtering edges by time) or aggregating data. We provided respective Python scripts and assisted participants with applying them to their data. Activities included loading the finalized data tables into the Vistorian using the import wizard.

Week 4: Network Exploration Using Node-Link Diagrams introduced node-link diagrams and related interactive features. We presented different exploration tasks [44] in the form of questions, e.g., *What to observe about network structure?*, *What does that mean for the data?*, or to interactively explore a network's change through time.

Week 5: Network Exploration Using Adjacency Matrices introduced adjacency matrices. We explained the visual encodings and how the node ordering is aiming to optimize for visual patterns in the matrix.

Week 6: Timeline, Maps, and Coordinated Views introduced timeline visualization and the geographic network visualization techniques, as well as the coordinated view feature in the Vistorian.

Again, throughout the course, participants were guided to prepare, visualize and explore their own data using the different network visualization features provided by the Vistorian. We also encouraged participants to share and discuss insights that they had gained and challenges they had encountered. Our course website [4] provides more details on the course and related materials.

5.3 Participants

36 participants from a wide variety of backgrounds, including history, social sciences, business & management, health & medicine, law & politics, natural sciences, and technology, education and event management registered for the course. In a pre-course questionnaire 18 (50%) participants stated that they had no experience with network exploration, 12 (33.3%) had explored one or two network visualizations prior to the course, and 6 (16.6%) participants stated experience with three or more network visualizations prior to the course. Due to the nature of the course, which was organized during the COVID-19 period and allowed participants to attend sessions synchronously online or watch recorded lectures asynchronously, we cannot report how many participants attended all sessions of the course. Of those who registered, 24 attended the first session synchronously, with subsequent course sessions being attended by 14 participants on average. Our findings are based on data collected across all participants, regardless of their pattern of course attendance.

5.4 Data Collection & Analysis

Our findings are based on the following data collected across the course.

Questionnaires. We asked all participants to fill out questionnaires prior to and after the course to capture their starting points and expectations when entering and their insights and reflections after the course. We received completed pre-questionnaires from 36 participants, and exit questionnaires from 8 participants.

Interviews. We invited all 36 course participants to a semi-structured interview where we asked them about their initial motivation to take the course, their lessons learned, to what extent they managed to create visualizations with the Vistorian, what strategies they applied, and what they found the most challenging about the course and the visualization process. To learn more about the Vistorian, we also included questions

about which features participants most appreciated, which ones they found most difficult to use, what additional features they would like to see added, and how they would overall describe their workflow with the Vistorian (See the supplementary material for the list of interview questions). 13 participants followed our invitation to be interviewed. All interviews were automatically transcribed prior to their analysis.

Course Recordings & Online Discussions. All course sessions were recorded, and we took notes of comments, questions, and discussions that drove the tutorial and drop-in sessions. We also captured inquiries and discussions on the course's Slack channel, where participants would ask questions about the course content and ask for help on their network visualization processes and challenges. Course recordings were not fully transcribed, but reviewed to complement our written notes.

Our analysis across this variety of data largely follows a thematic analysis approach [21, 34] where one author qualitatively coded participants' statements and the emerging coding was then discussed and iteratively refined between a team of three co-authors. In a first qualitative coding pass, we captured themes visible in our data in a top-down manner, driven by our research questions, and questions highlighted in our questionnaires and interviews. In a second coding pass, we allowed themes to emerge directly from our data collected. Our thematic analysis started already during the course, where we reflected on observations noted as part of course- and drop-in sessions. Emergent themes (e.g., observations around approaches to data preparation, how participants went about defining their first network, and approaches to visual exploration) were refined after the course, also in the light of participants' statements from the exit questionnaires and interviews.

Our findings from Study 2 are structured into four sections. We first describe participants' goals and types of course journeys (Sect. 6). Based on this we describe the range of observations with regard to network preparation, visualization, and exploration phases that participants went through (Sect. 7, Fig. 1). From these observations we derived an 8-step-model of network exploration which helps to formalize and contextualize participants' exploration processes (from data preparation to deriving findings, Sect. 8) as well as the barriers that can hamper these. Finally, we report on these barriers that some participants faced as part of these processes (Sect. 9).

6 PARTICIPANTS' GOALS & COURSE JOURNEYS

We start our findings by characterizing participants' starting points when entering the course, including their types of goals and questions (G1–G3) with respect to their data and analysis. We then provide a high-level overview of participants' learning journeys (J1–J3) through the course, illustrated by 14 representative participant profiles. For the remainder of the paper we refer to individual participants by the topical focus of their data (e.g., *P-Business*).

6.1 Initial Participant Goals

Statements from the pre-course questionnaires show that participants' motivations for engaging in network visualization varied in terms of the specificity of goals, exploration questions and the data at hand.

G1: Learning About Network Visualization To Inform Data Collection. Approximately 22.2% (8/36) of our course participants were still in the process of collecting data when entering the course or planned to start with the collection. Their main goal was to learn about network visualization and analysis in order to inform their data collection processes and questions of potential interest: *"I am preparing for network data collection but would like to see what can be analyzed before deciding on types of network data to collect."*[*P-SupplyChain*].

G2: Exploring Data to Help Define Questions. 50% (18/36) of our participants already had data, but no specific exploration questions prior to the course; Their goal was to *visualize and explore* their data to help generate questions (e.g., *P-Dance, P-Laureates*), check if network visualization applicable on their data (e.g., *P-RocksStudies*), or aimed to explore the quality of their data (e.g., *P-Documents*).

G3: Specific Questions To Inform Exploration. Approximately 27.8% (10/36) of our participants had defined questions about their data that they wanted to explore, for example, how a particular network had changed over time (e.g., *P-Business*), understanding the structure

of a network (e.g., *P-Criminology, P-Business*), or who talks to whom in a social network (e.g., *P-Carers, P-ElderlyCare*).

6.2 Participant Course Journeys

While all active participants (24/36) were able to create at least one network, they followed a range of course journeys, influenced by their data and initial goals. In order to anchor our findings on network exploration strategies and potentially hampering barriers, we illustrate the different types of journeys based on selected participant profiles.

J1: Vistorian Exploration using Demo Data. Participants who did not bring their own data to the course explored the features of the Vistorian based on the demo datasets provided, to learn about network visualization (G1). For example, *P-UrbanStudies* an engineer with a social sciences background and experience in network visualization was *"interested in learning about timeline and geographic map networks, plotting adjacency matrices, and learning what Vistorian is about."*

J2: Assisted Network Exploration. 2/24 participants required assistance in creating their first visualization (e.g. *P-Carers* health sector; no network vis experience). They struggled to define a network structure relevant to their questions. However, by the end of the course, they were able to create and explore network visualizations using their own data. Other participants, managed to create a first visualization on their own, but required assistance with refining their network structure to align with their goals and questions. For example, *P-Letters* (history; some experience in network vis) focused on data on a collection of historical documents and letters. Their goal was to understand the structure of the network spanned by documents and correspondents (G3). They—with the help of the course instructors—iterated on their initial network visualization leading to an adequate network visualization that enable interrogating the data according to these goals. *"before I explored my network, I supposed that the institution [X] was the most influential actor in my network, but it was [Y]."*[*P-Letters*].

J3: Independent Network Exploration. At least 5/24 participants actively and without assistance explored their data from the different perspectives offered by the Vistorian. For example, *P-Business*(G3) (computer science, social sciences, humanities; experience in network vis but not with the Vistorian) wanted to explore how funding rounds of commercial businesses assist in the growth of industries and business networks, and how the people behind businesses are linked. To explore these questions, they created and iteratively refined a number of temporal and geographic network visualizations and adjacency matrices. *P-RocksStudies* (G2) (Natural Sciences; some experience in network visualization) had a dataset describes studies and methodologies about a specific phenomenon related to rocks. They were able to create visualization but found it does not contribute to their research. *P-Documents*(G2) started with no research question in mind but used network visualization to figure out the best way to link nodes that might reveal interesting patterns in their data.

7 OBSERVATIONS

Based on our observations throughout the course and our analysis of the post-course questionnaires and interviews, we identified different aspects that influenced (1) how participants went about defining network structures, (2) how they iterated on their visualizations, and (3) how they approached their visual explorations of created networks. We describe our observations (O1–O10) below.

7.1 Defining Network Structures

The first part of the course (Weeks 1–3) focused on helping participants prepare and shape their data to create a first network visualization. Some participants with prior experience in network visualization were able to accomplish this easily (**J3**) and were able to skip ahead to the visual exploration phases. For participants with less experience, we scaffolded this process through lectures, activities (see Sect. 5.2), and individual assistance (**J2**). Participants without their own data (**J1**) did not have this problem since data was already defined and formatted. Below, we report on the factors that influenced participants' process of creating a first network visualization and the challenges we observed.

7.1.1 Preconceived Ideas and Mental Images

Our conversations with participants through the course revealed that around 9 participants had an initial mental image of the network they wanted to visualize in the Vistorian. These mental images were only partially influenced by participants' initial questions and goals but, rather, by their 'intuitive' understanding of the nature of their data and pre-conceived ideas of what their network *should* look like and what entities (nodes) and relationships (links) it expressed.

☉ **O1: Preconceived Ideas About Networks Can Blind Analysts to Alternative Approaches.** Around 2 participants thought of network visualizations predominantly as social networks showing people and their social relationships. Consequently, they initially defined nodes in their network as people and links as their respective relationships. It can blind participants to alternatives, and hindered them from exploring other mappings of nodes and links that could express different information about their data or support more specific goals. This reflects findings about *latent abstractions* [15], i.e., participants not having clear abstractions of their data, such as networks, tables or timelines. For example, *P-Carers* collected data about social relations between 80 patients and their carers from interviews. Initially, *P-Carers* wanted to create an ego-network centered around each patient. However, discussion and initial network visualizations revealed that the patients were not actually that relevant to the network, as *P-Carers* was interested in the *types* of carers and their relations through the act of caring. In other words, the collection of individual ego-networks had to be re-imagined as a network of relationships between roles of carers. Re-thinking her network structure, *P-Carers* reformatted her data tables several times again, a process common to data workers [8]. Similarly, *P-Letters's* data focused on "two persons exchanging letters on several topics." Similar to *P-Carers*, the participant imagined a network of authors (nodes), connected through multiple links, one for each letter. The initial network visualization—a two-node network with approx. 50 connections—was not useful. During discussion (J2), we suggested representing topics mentioned in the letters as nodes beside persons and link them through response type; this drastically changed the network structure and the questions that could be explored.

☉ **O2: Analysts have Mental Images of their Networks.** Besides preconceived ideas of what the nodes and links of a network should be, several participants had specific visual mental images of what their network would look like and what they would be able to observe: "I'm envisioning one map [(network)] per person with an interconnected web surrounding all the people, things, artifacts, systems, which make up care in their everyday life." [*P-ElderlyCare*]. Often, these images were disjoint from participants' goals and questions but, rather, were based on an intuitive understanding of their data. Similarly, *P-Documents* worked with labeled documents and wanted to create a network visualization based on the co-occurrence of labels for any two documents (nodes). The participant expected that this would result in clear network structure, however, their first visualization turned out to be cluttered to be readable: "I assumed the visualization will have the answer ready, but what I found it needed multiple rounds of data formation and understanding the underlying aspects to achieve it". This effect is commonly observed in visualization design processes [75].

7.1.2 Defining and Formatting Networks

40% of our participants (in particular those with less specific questions or goals) were initially unsure about the aspects in their data that should form the nodes and links in their network(s). When asked about particular challenges they faced during the course, one participant highlighted the difficulty of defining the structure of their network data: "deciding which and how information should be included prior to exploration [before having seen the network visualized]."

During the course, we facilitated this abstraction process using sketching exercises that asked participants to envision possible node and link (Sect. 5.2, Weeks 1&2). To a certain extent, this approach was successful enabling active participants to create at least one visualization of their dataset. However, besides defining nodes and links in ways that led to networks that were difficult to explore or simply misaligned

with participants' goals, we observed participants struggling to decide on particular aspects when it came to defining their network schema.

☉ **O3: Alternative Mappings to Nodes and Links.** Intuitively defining nodes and links often resulted in huge networks with over 10,000 nodes or links, turning most of the visualizations into colored hairballs. One strategy discussed in the course was to reduce network size by filtering by node or link-type, filtering nodes based on their degrees only, or slicing the network along time (Sect. 5.2, Week 3). Another common strategy is to transform the network, through graph transformations such as projections or clustering. Especially link projection was new to participants and helped improve their network models in the case of multi-modal networks. For example, *P-Dance* tried to identify participants (nodes) who participated together in one or more events, whereas the initial data they had was who participated in which event. *P-Archaeo* tried to explore the movement of artifacts from one site to another over time and which artifacts were found at the same sites. Thus, we assisted them in creating the respective network projection. The resulting network enabled a meaningful cluster analysis.

P-Festivals's geographic network described links between countries and specific festival locations within the same single city. The respective network featured countries on the one side and street-level information in a single city on the other, making a good scaled representation of those locations impossible. Through discussion, we explored three alternative networks. Network 1 focused on that single city, with nodes being festival locations and links showing that two locations hosted artists from the same country. Network 2 would keep countries as nodes but collapse all locations within that single country into a single node to result in a more readable map visualization. Network 3 would be the initial network but visualized as node-link or matrix only. Without discussing such aspects with the participants, some workarounds might be hard for them to get by their own.

☉ **O4: Dealing with Multiple Occurrences of the Same Data Instance.** Several participants had data where the same instance of a data points occurred multiple times (e.g., because of contextual or temporal changes). These participants found it difficult to decide whether to represent each data points as a separate node, or perform aggregation and represent all data points from the same entity as a single node. For example, *P-Carers* had multiple instances of the same family doctors in their data as these doctors were connected to several patients. Having different node instances representing each doctors made it hard to identify roles of high importance. They, finally decided to use a single node to represent each carer role, e.g., 'family doctor'.

Similarly, *P-Laureates*, working with a dataset of laureates across multiple years, aimed to find those who shared the same types of awards. They first decided to represent awards in each year as individual nodes. However, it turned out that this offered a limited understanding of laureates across years. In a subsequent iteration, *P-Laureates* therefore represented each award title (regardless of the year) as a single node. Creating one node per year and linking to the laureates in that years would help *P-Laureates* answer their question about when the same person had received the award twice.

☉ **O5: Defining Link Types & Weight.** Another common issue was the definition of link types and how to map relevant data to link weight. One participant mentioned that "deciding the type of link between the nodes [...] which aspect of the connection [...] is most relevant to answering the research question" (anonymous) was the most challenging aspect of their network visualization process.

Particular problems arose from wanting to assign link weights based on non-quantitative attributes, such as rankings expressed through words (e.g. 'strong', 'weak', *P-Carers*) or different types of activities ('applied', 'interviewed', 'job-offered'). *P-Carers* initially used a ranked scale from 1–10 (1 representing the most important links). However, in the Vistorian low numbers assigned to link weight result in faint lines, while link weights ranked higher are visually emphasized. Furthermore, they found that the distribution of values in their data was rather small, making it difficult to visually differentiate between important and less important links. While these issues can be fixed quickly and might not be specific to network visualizations, again, they provide insights into the small barriers that people might encounter and

which eventually can prevent them from successfully using a tool.

🕒 **O6: Formatting Data into Node and Link Tables.** In line with other network visualization tools, the Vistorian supports node and link tables as data formats. In line with previous work [35], many participants were initially confused what type of table formatting to apply. Further confusion resulted from declaring links with changing weight over time, which required an individual row in the link table for each time point (and their weight), linked together by an ID column to identify identical links. Without IDs, the Vistorian would interpret each row as a different link. On the other side, this option could be useful to show all possible link weights at the same time, i.e., through individual links between the same two nodes. Most participants were able to determine the suitable format for their network.

7.2 Iterating on the First Visualization

As illustrated above in several examples, most participants iterated on their first visualization multiple times. We observed participants—both those with and without prior experience in network visualization—not only re-defining their nodes and links but also their data, based on issues they encountered with their initial visualization.

🕒 **O7: Data-level Refinement.** One common issue that participants encountered with their initial network visualization (e.g., *P-Festivals*, *P-Laureates*, and *P-Documents*) was that too many nodes or edges made the network appear as a hairball that was too cluttered to explore in a meaningful way. The course covered strategies such as slicing the network along time or filtering nodes and edges based on relevance or type. As illustrated above, several participants had defined network schema that resulted in duplicate nodes or links—either because of issues in the data, or because of the very nature of the data. For example, *P-Festivals* found that the same venue recruited certain ensembles multiple times for different events. *P-Documents*'s data included labeled documents that included multiple co-occurrence links with other documents. During the course, we provided technical support (in the form of dedicated Python scripts) to help participants explore duplicate links and, based on their goals and questions, transform these meaningfully (e.g., aggregating data points to be represented in the form of link weight) (Sect. 5.2, Week 3). This helped several participants reduce the size of their networks and highlight important patterns in their data without affecting the nature of the network.

7.3 Visual Exploration

🕒 **O8: Tool Features Can Inspire Exploration Strategies.** Interactive or visual features that seem trivial to visualization designers and exploration experts can be a game changer to novice analysts, such as time sliders or the timeline visualizations (“*The time-slider provided an entirely different perspective on my data.*”). They can provide important impulses for the exploration of data and inform analysis questions: “*I largely ignored the time element of my data up to this point. The timeline provided an entirely new dimension to my data and certainly changed my perspective.*”. Features such as the timeline assist in finding the story the analyst would like to tell.

Similarly, analysts appreciated the coordinated views showing node-link and matrix representation at the same time. Participants liked how they enabled interpreting and comparing different visualizations of the same network at the same time and how matrix and node-link diagrams complemented each other (“*Even if I missed noticing that there are two groups in the node-link diagram, the matrix would show me that.*”[*P-Criminology*]).

🕒 **O9: Unfamiliar Visual Encodings are not immediately understood.** A well known problem, yet most visualizations in the Vistorian were well understood by course participants. Adjacency matrices caused some confusion in the beginning. While understanding the matrix ordering was easier understood than we thought, most questions arose from the visual encoding of cells in the matrix, particularly when cells used a multitude of visual encodings: multiple links (cell split into equally sized rectangles), link type (color), link weight (opacity), and link direction (gradient)

After understanding the visual encoding, the challenge participants faced was to interpret their own data using these encodings. This

observation holds for both the layout of nodes in a node-link diagram and the ordering in a matrix. Participants reported the need to develop trust in the visualizations and their visual observations.

🕒 **O10: Building Trust In Visualization.** Participants often asked us about measuring the reliability of the observations in the visualization and building trust with respect to the visualization: “*How do I assess whether what I’m visualizing is meaningful, useful, reliable from the data?*”[*P-Carers*]. Moreover, the issue was about the general communication of findings and the scientific rigor in visualization, i.e., to understand “*reliability in network science*”[*P-Carers*] and “*where [visualization] fits on the spectrum of qualitative, quantitative and also mixed methods. Partly, because the course talked a bit about how analysts can always cut things multiple ways, and [then] need to know what is the meaning of [their findings].*”[*P-Carers*]. These questions had implications about the participants’ own work and how much they could trust their own observations: “*[Hence,] how will I know that something matters or has meaning, rather than just it’s a sort of artifact of having played around with it for a long time.*”[*P-Carers*].

8 STEPS IN NETWORK EXPLORATION

Our observations point to specific barriers in the *process* of visual network exploration. To better contextualize these barriers, this section develops a model of 8 steps that we found in exploring networks. The model is informed by our observations (O1-O10) and the participant journeys (J1-J3). It is also inspired by observations in visualization literacy more generally, notably the need to translate abstract domain concepts into visual encodings and back, in order to solve tasks with visualizations [20]. It also links to Norman’s *Gulfs of Execution and Evaluation* for cognitive processes that accompany interactions in interactive systems [55]. Eventually, there are parallels to Munzner’s Nested Model describing the increasing abstraction and formalization from domain concepts and problems into algorithms and systems [52].

In a first phase (left side in Fig. 1), data is formalized from abstract domain concepts (i.e., ideas) into visual encodings (i.e., concrete forms on a display) in a network visualization. Then, based on this visualization, information is interpreted to inform findings (right side in Fig. 1) leading back from forms in the visualization to findings in the domain. Throughout these two phases, we can describe 8 steps across an analysts’ journey from *domain* to *network* to *data* to *visualization* and back. **1) Identify domain concepts** in the data (e.g., people, events, locations, transactions, participation in events). This is a necessary preparatory step; **2) Define a network** to turn domain concepts into network concepts, i.e., nodes, links, link weight, direction and (if necessary) apply network transformations; **3) Format data** to turn the network concepts into a machine readable format such as node and link tables; **4) Import data and specify data schema** to create mappings from the data format (i.e., columns in the case of the Vistorian) to semantics and visual variables in a visualization; **5) Interactively explore visualization(s)** through interaction and choosing among different visualizations; **6) Perceive visual patterns** in a visualization, understanding meaningful constellations of pixels, colors, etc; **7) Interpret visual patterns** to understand visual patterns as network concepts, i.e., ‘a cluster’ or ‘a central node’; **8) Interpret network concepts**, i.e., clusters, paths, central nodes in the respective domain and understand their meaning for that domain and respective questions.

Depending on an analyst’s expertise and the specific tool used, these steps can be less deliberate and more based on tacit knowledge. For example, an experienced analyst may directly format their network structure into a machine readable format as they define it, or re-work an existing node table to tailor it to a new question. The fact that data can be already formatted *before* the analyst decided on the network structures does not invalidate the model but simply means that a data format is required. Similarly, different tools can support different steps of the process. However, the model is deliberately kept simple to picture the decisions involved in network exploration and to inform future support (tools and methods) to overcome the barriers.

9 BARRIERS TO NETWORK EXPLORATION

From our 24 active participants, nearly all succeeded in creating at least one network visualization of their own data using the Vistorian. One participant's data (*P-RocksStudies*) turned out not to be useful to be mapped as a network. 11/24 participants used the Vistorian with little help, and 5 participants did not experience any barriers at all. Each of the other participants experienced between 1-5 barriers each, with the specific barriers experienced greatly varying across participants. Through discussion, and explanations, we were able to resolve most if not all of these barriers. In the following, we now describe 8 barriers (numbers in brackets indicate number of participants facing this barrier) and discuss how we mitigated them in our course.

¶¶¶ B1: Missing Goals & Questions (3)—Missing goals for explorations are not a barrier *per se*. A more open-ended approach can inspire creative perspectives on the data and novel angles of exploration. However, a lack of specific goals can become a barrier to all steps of network exploration, in particular to early tasks, i.e., Steps 1 and 2 in our model, where relevant data aspects have to be identified and translated into a network schema, and to Steps 4 and 5 where decisions about suitable network visualization techniques influence what types of patterns can be explored. In the worst case, a lack of goals and questions can lead to irrelevant findings or incorrect conclusions, e.g., via drill-down fallacies [45]). *Mitigation strategies*—Use constructive visualization methods [40] (e.g., sketching) to explore data in a low-cost and open-ended way that helps identify points of interest, goals and questions. Show example network visualizations and explaining what they can reveal (**O8**), to stimulate creative thinking about network visualization and how it can be applied to facilitate sense-making.

¶¶¶ B2: Pre-conceived Ideas & Mental Images (5)—Like missing goals, we think that mental images (which can also be influenced by preconceived ideas about network visualization) about what the data might look like as a network (**O2**) are an essential part of visualization and exploration in that they can inform goals and starting points to network exploration. However, both preconceived ideas and mental images can turn into barriers. They can hamper tasks in Steps 2 & 3, leading to a mapping of domain concepts to network structures that may not match with analysts' goals and questions (**O3**). Even worse, preconceived ideas and mental images can hinder a creative exploration process that allows for new perspectives on the data. For example, they may lead to analysts dismissing new angles of exploration that may or may not include unfamiliar visual encodings (**O9**), to preventing analysts from exploring alternative mappings of domain concepts to nodes and links (see *P-Letters*). In the worse case, preconceived ideas and mental images can negatively influence the exploration and scrutiny of the data all together, leading to confirmation bias, although we have not observed this in our course. *Mitigation strategies*—An early visualization where analysts can see their ideas realized can be a pivotal point to help iterate on early steps of the data formation process or to explore alternative visualization techniques. Support the externalization (e.g., via sketching) and interrogation (e.g., via probing questions) of ideas and mental images early on in the network visualization process in a lightweight manner to help analysts align their visions with their goals and questions. Approaches that allow a critical reflection on data formation and design processes could facilitate this.

¶¶¶ B3: Deciding on a Network Structure (4)—Defining network structures such as nodes, links, and link weight has a profound impact on all the consequent steps, as illustrated in Fig. 1. Depending on an analyst's experience or domain, these mappings can be trivial. However, we found, in particular, participants with little experience in network visualization, struggled with deciding on a network structure (tasks related to Steps 2 & 3 of the network exploration process), which created a profound barrier and delayed subsequent processes. We added an extra session on data preparation into our course to address this. *Mitigation strategies*—Breaking down the process of defining a network structure into different activities can help overcome this barrier (e.g., identifying potential nodes and links on paper, sketch potential outcome visualizations based on this, and formalize outcomes of these processes as concept maps; see our course activities in Sect. 5.2). However, there are still open questions about how such activities can

be integrated into network visualization tools.

¶¶¶ B4: Choosing The Right Level of Abstraction (6)—The initial network visualizations of participants not always yield insights, because an intuitive mapping of domain concepts to nodes and links led to visually cluttered networks due to too many nodes or edges (or both). Participants struggled with introducing the right level of abstraction to create more meaningful networks. We define abstraction as the process of deducing a new network from the initial network that describes the entire dataset. This notion of abstraction is different from Bigelow's et al. [15] which refers to data abstractions as tables, networks, timelines or alike. *Mitigation Strategies*—In the course, we discussed options and gave participants access to dedicated Python scripts to facilitate network size reduction, as well as transformation and aggregation strategies. Tool support in this area exists (e.g., [14]), and, again, sketching can help identify goals and approaches to abstraction.

¶¶¶ B5: Choosing the Appropriate Data Formats (2)—Choosing the appropriate (machine readable) data format and transforming data accordingly (Step 3) can become a barrier through a range of issues (**O6**), e.g., (1) not understanding the data format required, (2) confusing different formatting options, and (3) inconsistent formatting. Besides general benefits of tables for making data accessible [8], node and link tables are relatively straightforward to understand, are editable in common applications such as Excel, and are successfully used in other network visualization tools [71]. However, some participants struggled to decide what information to include in a node table and what would be better represented as a link table. *Mitigation strategies*—Provide illustrated guides for table formatting (e.g., [5]); promote individual hands-on activities to table formatting. At a tool-level, it is important to scaffold data import in a step-by-step manner, remind analysts how to interpret different tables and provide side-by-side views of data tables and the resulting visualization (e.g., as already done in NodeXL [71]). Data validation and consistency checking can be done using different tools (e.g., Python).

¶¶¶ B6: Importing Data and Defining Schemata (2)—A data schema creates an explicit mapping between the data and the visualization. For example, in an earlier version of the Vistorian, users uploaded a link table and then had to manually specify the role of each column. In addition, they could upload a node table and a location table, or click a button to retrieve geo-coordinates for locations mentioned in the tables. Observations from early workshops showed that planning and coordinating all these actions is a barrier for some analysts (e.g. *P-Migrants*). Our import wizard (Sect. 5.1) addresses this issue, but can slow down the data import for advanced analysts. *Mitigation strategies*—Scaffold data import and mapping processes through wizards with detailed explanations, while allowing advanced analysts to streamline the process.

¶¶¶ B7: Interpreting Visual Patterns in Visualization (2)—Visual patterns in network visualizations need to be understood as network concepts before they can be properly interpreted (Steps 6 & 7). This includes not only low-level encodings of visual elements (e.g., through size or color), but their composition into more complex visual constructs. For example, blocks of cells in an adjacency matrix represent a densely connected cluster, while the same cluster may appear as a set of overlapping arcs in a time arc visualization. While we observed a steep learning curve in reading and interpreting adjacency matrices (**O9**), understanding the richness of network visualization can pose barriers. For example, it can be overwhelming to engage with a network visualization for the first time and to explore it for patterns and clusters—whether or not the technique is familiar. Also, to add complexity, interaction can lead to changes of visual patterns on-the-fly. To better understand visual patterns, participants reported on multiple coordinated views helping to interpret and verify a pattern visible in one visualization through another one (**O8**). *Mitigation strategies*—Document and illustrate network visualization techniques through demos and practical use cases; provide hands-on tutorials exploring different network visualizations and illustrate their strengths and limitations, e.g., through visualization cheat sheets [76]; support multiple coordinated views.

¶¶¶ B8: Establish Trust in a Network Visualization (1)—Trust in visualization is a generic topic [50] and poses a potential barrier to network visualization in itself (**O10**). It can influence if analysts engage in

a network visualization process in the first place, how findings are derived from the visualizations, and if a network visualization is deemed valid for evidence in (scientific) communication. One problem can be unfamiliar visualization (e.g., adjacency matrix) and their respective construction methods such as ordering (in the case of matrices), or layouts in node-link diagrams. Another problem is understanding the *provenance*, including the many decisions along the network creation: conceptualization of nodes and links, network transformations, and filtered elements. For network visualization, there seems to be less of an established culture for how to create trustworthy visualizations in the first place and use them for scientific inquiry. *Mitigation strategies*—Provide sufficient explanation and documentation of the strengths and limitations of individual visualization techniques (including interpretation pitfalls); explain underlying algorithm, their assumptions, and possible artifacts they can introduce [76]; show examples of anti-patterns where a visualization is misleading or introduces artifacts (e.g., different layout algorithms, clutter in node-link diagrams); use examples of visualizations to show limitations of analytic approaches to network science in terms of network metrics [23,49] (e.g., an algorithm identifying clusters not visible in a visualization [24]); show examples of network visualization success stories and uses in journalism [19].

10 DISCUSSION

We described 8 potential barriers that pose problems to *some* analysts during *some* of their exploration process. In our course, these barriers could all be solved through discussion, rather than being built-in problems of specific tools or due to human incapacity. We think these barriers are due to inexperience and underestimating the decisions involved in sensemaking with data and visualization. Eventually, analysts will learn how to navigate the barriers but, if left on their own, our log study and discussions suggest that these barriers indeed block analysts from progressing and can lead to resignation. Like fire-drills, these barriers highlight the importance of training, cautioning analysts upfront, and thinking of better methods and tool support.

In summary, our findings also highlight the data preparation stages of network visualization as an important source for insights in themselves and how this process helps novice analysts learning: “*Getting the data prepared to analyze visually [. . .] helped me to realize things about the data that I did not know. It was less about the actual visualizations at the end, it was a lot more about looking at the types of structures we had and thinking about the process...*”[P-Business].

Implications for Education—Our work network visualization teaching approaches and related questions to be addressed in the future:

- **Work with Participants’ Data.** Allowing participants to work with their own data was hugely important for the success of our course. At the same time, demo data sets are important to explain and illustrate particular aspects of network visualization. Ideally, demo data should be relevant to the course audience and consistent across the course to allow for the comparison of techniques.
- **Focus on Understanding Goals.** Early course activities should focus on helping participants to reflect on their own goals, which is also beneficial for course instructors to provide individual support. Pre-course questionnaires can be beneficial here.
- **Provide Activities to Identify & Mitigate Barriers.** We found that sketching activities provide an easy entry point for analysts to engage in the data preparation stages of network visualization. Additional research is needed to explore how to frame sketching activities in this context and also think beyond sketching.
- **Balance Independent Exploration & Guidance.** We tried to strike a balance between scaffolding participants’ explorations and allowing them to come up with and explore their own strategies.
- **Define Clear Learning Goals and Essential Skills.** This is an area to explore more in the future. Our barriers can serve as waypoints and benchmarks for these learning goals.
- **Create Teaching Resources.** We need to start exploring additional resources that can facilitate teaching network visualization in ways that mitigate the barriers discussed in this paper. For example, success stories and galleries of successful network visualizations (beyond node-link diagrams) can help to show what is possible with network

visualization and can motivate novices and expert analysts to consider network visualization as a method in the first place.

- **Help Establish Trust.** To build more trust in network visualizations, future research could look into explaining network visualization algorithms in concise and accessible ways (e.g., cheatsheets [76]).

Implications for Tool Design—Our work shows that the range of data and research goals that analysts bring to network visualization is vast. It is therefore sensible to start thinking about network visualization tools and teaching that consider the process of building visualization from the analyst’s perspective and allow for different entry points in terms of exploration goals, data, visualization techniques and exploration strategies. Below are some design considerations that can enhance the analyst’s experience in visualizing and exploring data when designing tools as well as engaging with network analysts.

- Support the analyst’s perception of the data form by offering a concept-demonstrating miniature (e.g., **concept map**) or diagram, an approach already taken by Origraph [14]. Changes and selections to this concept map can be reflected in the network.
- Encourage **incremental network building** and iterative exploration; this is similar to the idea of enabling ‘progressive evaluation’ in the Cognitive Dimensions of Notations framework [32].
- Provide **coordinated multiple views** using different types of visualization. This not only allow the analysts to take advantage of their complementary strengths, but also allows them to learn how to interpret an unfamiliar visualization by observing how it changes as they interact with a more familiar one.
- Support **creative approaches** to creating networks. For example, show possible networks derived from a given a “seed” network through aggregation, projection, and filtering.
- **Explain visual patterns and encoding as part the tool**, e.g., using automatic annotations suggesting explanations for specific patterns or miniature galleries of visual patterns found inside a dataset.
- **Automatically recommend or highlight** potentially insightful views and data pattern.

Methodology, Limitations, and Open Questions—In-the-wild research is inherently challenging [17, 70]. Our findings are situated in the specifics of our course and course participants, as well as the tool we used. Since analysts may have enrolled because they had particular conceptual challenges in network visualization, we cannot say how prominent our observed barriers are in other settings (e.g., among analysts with different backgrounds and training). Similarly, those participants who engaged in discussions and interviews might not be those who fail *silently* for reasons we were not able to capture. However, the consistency of our barriers across participants hints towards general issues. In the future, we need better methods to collect data about analysts in the wild, especially, to capture (thinking) processes where analysts are not using a computer where actions can be logged. Our experience with mini-questionnaires [69] and microentries [16] were not encouraging due to missing engagement by users. Participants’ workflows were influenced by the Vistorian, its required data formats, import routines, and visualizations. Different tools may impact the existence and gravity of barriers. Still, many of our observations are independent of the particular tool, including the barriers during the *definition* phase (e.g., **B1, B2, B3**).

11 CONCLUSIONS

We have shown that some of the reasons for people having limited engagement with network visualization tools are barriers in preparing their data and planning their exploration, such as being unable to transform what they have in mind into compatible data structures. Throughout this paper we have described such issues in detail, and discussed how the data formation process is an essential unit of the visualization tool’s design. These findings *must not* be used to blame the analyst, but rather shed light on the complex nature of understanding our world through data and visualization. We hope our paper contributes to the growing body of literature in visualization and data literacy, as well as visualization pedagogy. We hope our research will inspire similar research for other data types and domains and lead to a greater awareness of people’s thinking process during dealing with abstract data.

REFERENCES

- [1] B. Alper, B. Bach, N. Henry Riche, T. Isenberg, and J.-D. Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 483–492, 2013.
- [2] D. Archambault, H. Purchase, and B. Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552, 2010.
- [3] B. Bach, M. AlKadi, J.-D. Fekete, C. Plaisant, V. Serrano, N. H. Riche, and N. Dufournaud. The vistorian. <https://vistorian.github.io/>. Accessed on July, 1st, 2022.
- [4] B. Bach, M. AlKadi, U. Hinrichs, and J. Scott-Brown. Free 6-Weeks Online Course on Interactive Visual Network Exploration. <https://vistorian.github.io/courses.html>. Accessed on July, 1st, 2022.
- [5] B. Bach, M. AlKadi, U. Hinrichs, and J. Scott-Brown. Vistorian Data Formats Course. <https://vistorian.github.io/formattingdata.html>. Accessed on July, 1st, 2022.
- [6] B. Bach, E. Pietriga, and J.-D. Fekete. Graphdiaries: Animated transitions and temporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2013.
- [7] B. Bach, N. H. Riche, R. Fernandez, E. Giannidakis, B. Lee, and J.-D. Fekete. NetworkCube: bringing dynamic network visualizations to domain scientists. In *Posters of the Conference on Information Visualization (InfoVis)*, 2015.
- [8] L. Bartram, M. Correll, and M. Tory. Untidy data: The unreasonable effectiveness of tables. *arXiv preprint arXiv:2106.15005*, 2021.
- [9] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*, 2009.
- [10] A. Batch and N. Elmqvist. ori. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):278–287, 2017. doi: 10/gcp8bz
- [11] M. Baur. *visone - Software for the Analysis and Visualization of Social Networks*. PhD thesis, Karlsruhe Institute of Technology, 2008.
- [12] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. *EuroVis (STARs)*, 2014.
- [13] S. Bender-deMoll and D. A. McFarland. The art and science of dynamic network visualization. *Journal of Social Structure*, 7, 2006.
- [14] A. Bigelow, C. Nobre, M. Meyer, and A. Lex. Origraph: Interactive network wrangling. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, Oct. 2019. doi: 10/gpsvww
- [15] A. Bigelow, K. Williams, and K. E. Isaacs. Guidelines for pursuing and revealing data abstractions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1503–1513, 2020.
- [16] J. E. Block and E. D. Ragan. Micro-entries: Encouraging deeper evaluation of mental models over time for interactive data systems. In *Proceedings of the 2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 38–47, 2020. doi: 10/gpsvws
- [17] E. M. Bonsignore, C. Dunne, D. Rotman, M. Smith, T. Capone, D. L. Hansen, and B. Shneiderman. First steps to netviz nirvana: Evaluating social network analysis with NodeXL. In *Proceedings of the 2009 International Conference on Computational Science and Engineering*. IEEE, 2009. doi: 10/dx2p6n
- [18] K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016.
- [19] L. Bounegru, T. Venturini, J. Gray, and M. Jacomy. Narrating networks: Exploring the affordances of networks as storytelling devices in journalism. *Digital Journalism*, 5(6):699–730, 2017.
- [20] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, 2014.
- [21] R. E. Boyatzis. *Transforming Qualitative Information. Thematic Analysis and Code Development*. Sage Publishing, 1998.
- [22] C. Chang, B. Bach, T. Dwyer, and K. Marriott. Evaluating perceptually complementary views for network exploration tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1397–1407, 2017.
- [23] H. Chen, U. Soni, Y. Lu, R. Maciejewski, and S. Kobourov. Same stats, different graphs. In *International Symposium on Graph Drawing and Network Visualization*, pp. 463–477. Springer, 2018.
- [24] K.-T. Chen, T. Dwyer, B. Bach, and K. Marriott. It’s a wrap: Toroidal wrapping of network visualisations supports cluster understanding tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.
- [25] C. Cramer, L. Sheetz, H. Sayama, P. Trunfio, H. E. Stanley, and S. Uzzo. NetSci high: Bringing network science research to high schools. In *Studies in Computational Intelligence*, pp. 209–218. Springer International Publishing, 2015. doi: 10/gpb58v
- [26] Cyram Inc. Netminer - social network analysis software.
- [27] W. de Nooy, A. Mrvar, and V. Batagelj. *Exploratory social network analysis with Pajek*. Cambridge University Press, Cambridge, 2018.
- [28] O. ElTayeb and W. Dou. A survey on interaction log analysis for evaluating exploratory visualizations. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV)*, pp. 62–69, 2016.
- [29] J.-D. Fekete. Interaction Trace Manager. <https://github.com/INRIA/intertrace>. Accessed on July, 1st, 2022.
- [30] J.-D. Fekete. Reorder.js: A javascript library to reorder tables and networks. In *IEEE VIS 2015 Posters*, 2015.
- [31] M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, pp. 17–24, 2004. doi: 10.1109/INFVIS.2004.1
- [32] T. R. Green. Cognitive dimensions of notations. *People and computers V*, pp. 443–460, 1989.
- [33] C. Grimes, D. Tang, and D. Russell. Query logs alone are not enough. In *WWW 2007 Workshop on Query Log Analysis: Social and Technological Changes*, 2007.
- [34] G. Guest, K. M. MacQueen, and E. E. Namey. *Introduction to Applied Thematic Analysis*. Sage Publishing, 2012.
- [35] D. L. Hansen, D. Rotman, E. Bonsignore, N. Milic-Frayling, E. M. Rodrigues, M. Smith, and B. Shneiderman. Do you know the way to SNA?: A process model for analyzing and visualizing social media network data. In *Proceedings of the 2012 International Conference on Social Informatics*. IEEE, Dec. 2012. doi: 10/ggb6g3
- [36] J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. *Information Visualization*, 13(2):111–133, dec 2012. doi: 10/f5wggh
- [37] N. Henry and J.-D. Fekete. Matlink: Enhanced matrix visualization for analyzing social networks. In *Proceedings of the IFIP Conference on Human-Computer Interaction*, pp. 288–302. Springer, 2007.
- [38] D. Holten, P. Isenberg, J. J. van Wijk, and J.-D. Fekete. An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *Proceedings of the IEEE Pacific Visualization Symposium*. IEEE, Mar. 2011. doi: 10/cn3tm2
- [39] Y.-H. Hung and P. Parsons. Assessing user engagement in information visualization. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1708–1717, 2017.
- [40] S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauerer. Constructive visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS ’14, p. 433–442. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10/gfspd2
- [41] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, Sept. 2011. doi: 10/bs2nc2
- [42] S. Kasica, C. Berret, and T. Munzner. Table scraps: an actionable framework for multi-table data wrangling from an artifact study of computational journalism. *IEEE Transactions on visualization and computer graphics*, 27(2):957–966, 2020.
- [43] N. Kerracher, J. Kennedy, and K. Chalmers. A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 21(10):1160–1172, Oct. 2015. doi: 10/f7qbxq
- [44] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on Beyond time and errors novel evaluation methods for information visualization - BELIV '06*. ACM Press, 2006. doi: 10/b2pm7w
- [45] D. J. L. Lee, H. Dev, H. Hu, H. Elmeleegy, and A. G. Parameswaran. Avoiding drill-down fallacies with VisPilot: assisted exploration of data subsets. In *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI 2019, Marina del Rey, CA, USA, March 17-20, 2019*, pp. 186–196. ACM, 2019. doi: 10/gpq2rd

- [46] Z. Liu, S. B. Navathe, and J. T. Stasko. Ploceus: Modeling, visualizing, and analyzing tabular data as networks. *Information Visualization*, 13(1):59–89, jun 2013. doi: 10/f5t2qt
- [47] N. Mahyar, S.-H. Kim, and B. C. Kwon. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, vol. 3, p. 2, 2015.
- [48] Maptiler: Maps for developers. <https://www.maptiler.com/>. Accessed on July, 1st, 2022.
- [49] J. Matejka and G. Fitzmaurice. Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1290–1294, 2017.
- [50] E. Mayr, N. Hynek, S. Salisu, and F. Windhager. Trust in Information Visualization. In *Proceedings of the EuroVis Workshop on Trustworthy Visualization (TrustVis)*. The Eurographics Association, 2019. doi: 10.2312/trvis.20191187
- [51] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson. How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.
- [52] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.
- [53] C. Nobre, M. Meyer, M. Streit, and A. Lex. The state of the art in visualizing multivariate networks. In *Computer Graphics Forum*, vol. 38, pp. 807–832. Wiley Online Library, 2019.
- [54] C. Nobre, D. Wootton, L. Harrison, and A. Lex. Evaluating multivariate network visualization techniques using a validated design and crowdsourcing approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, 2020. doi: 10/gpsvwv
- [55] D. A. Norman. *The Design of Everyday Things*. Basic Books, Inc., USA, 2002.
- [56] NWB Team. Indiana University, Northeastern University, and University of Michigan. Network workbench tool., 2006.
- [57] M. Okoe, R. Jianu, and S. Kobourov. Revisited experimental comparison of node-link and matrix representations. In *Lecture Notes in Computer Science*, pp. 287–302. Springer International Publishing, 2018. doi: 10/gpsvwz
- [58] A. Perer and B. Shneiderman. Systematic yet flexible discovery. In *Proceedings of the 13th International Conference on Intelligent User Interfaces - IUI '08*. ACM Press, 2008. doi: 10/cpvvcz
- [59] A. Perer and B. Shneiderman. Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines. *IEEE Computer Graphics and Applications*, 29(3):39–51, may 2009. doi: 10/dh4fgc
- [60] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 109–116, 2004.
- [61] H. C. Purchase, R. F. Cohen, and M. I. James. An experimental study of the basis for graph drawing algorithms. *ACM Journal of Experimental Algorithmics*, 2:4, jan 1997. doi: 10/fkqn85
- [62] D. Ren, L. R. Marusich, J. O'Donovan, J. Z. Bakdash, J. A. Schaffer, D. N. Cassenti, S. E. Kase, H. E. Roy, W. yi (Sabrina) Lin, and T. Höllner. Understanding node-link and matrix visualizations of networks: A large-scale online experiment. *Network Science*, 7(2):242–264, June 2019. doi: 10/gpsvw2
- [63] M. Roughan and J. Tuke. Unravelling graph-exchange file formats, 2015. doi: 10.48550/ARXIV.1503.02781
- [64] S. P. Borgatti and M.G. Everett and L.C. Freeman (Analytic Technologies). UCINET for windows: Software for social network analysis., 2002.
- [65] H. Sayama, C. Cramer, M. A. Porter, L. Sheetz, and S. Uzzo. What are essential concepts about networks? *Journal of Complex Networks*, 4(3):457–474, 2016.
- [66] S. Schöffel, J. Schwank, and A. Ebert. A user study on multivariate edge visualizations for graph-based visual analysis tasks. In *Proceedings of the 2016 20th International Conference Information Visualisation (IV)*. IEEE, July 2016. doi: 10/gpsvw3
- [67] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and interacting with geospatial networks: A survey and design space. In *Computer Graphics Forum*, vol. 40, pp. 5–33. Wiley Online Library, 2021.
- [68] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. doi: 10/f4fv7x
- [69] V. Serrano Molinero, B. Bach, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Understanding the Use of The Vistorian: Complementing Logs with Context Mini-Questionnaires. In *Proceedings of the Visualization for the Digital Humanities Workshop*. Phoenix, United States, Oct. 2017.
- [70] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools. In *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06*. ACM Press, 2006. doi: 10/c6vfz6
- [71] M. Smith, A. Ceni, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, and C. Dunne. NodeXL: a free and open network overview, discovery and exploration add-in for excel 2007/2010/2013/2016 from the social media research foundation, 2010.
- [72] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave. Analyzing (social media) networks with NodeXL. In *Proceedings of the fourth international conference on Communities and technologies - C&T '09*. ACM Press, 2009. doi: 10/bbkvtj
- [73] Stanford University Humanities + Design lab. Palladio: Visualize complex historical data with ease.
- [74] A. Vogogias, D. Archambault, B. Bach, and J. Kennedy. Visual encodings for networks with multiple edge types. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 1–9, 2020.
- [75] J. Walny, C. Frisson, M. West, D. Kosminsky, S. Knudsen, S. Carpendale, and W. Willett. Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):12–22, 2019.
- [76] Z. Wang, L. Sundin, D. Murray-Rust, and B. Bach. Cheat sheets for data visualization techniques. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Apr. 2020. doi: 10/gnrq7t
- [77] J. wook Ahn, C. Plaisant, and B. Shneiderman. A task taxonomy for network evolution analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):365–376, Mar. 2014. doi: 10/f58vqt
- [78] V. Yoghourdjian, D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu. Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Visual Informatics*, 2(4):264–282, Dec. 2018. doi: 10/ggjn5
- [79] A. Zoss, A. Maltese, S. M. Uzzo, and K. Börner. Network visualization literacy: Novel approaches to measurement and instruction. In *Network Science In Education*, pp. 169–187. Springer International Publishing, 2018. doi: 10/gpsvw4