Journal of Perceptual Imaging® 8: 1–12, 2025.

Time-Dependent Route Planning on Maps: A User Study

Annika Bonerath[1], Claus-Christian Carbon[2], Silvia Miksch[3], Maurizio Patrignani[4], Alessandra Tappini[5]

[1] *Universität Bonn, Germany;* [2] *University of Bamberg, Germany;* [3] *TU Wien, Austria;* [4] *Università Roma Tre, Italy;*
[5] *Università degli Studi di Perugia, Italy*
E-mail: bonerath@igg.uni-bonn.de

Abstract. *Routing through a dynamic environment is mostly carried out by using maps that integrate information about time-dependent parameters, such as traffic conditions and spatial constraints, which is a challenging and cumbersome task. We address the complex scenario where a user has to plan a route on a network that is dynamic with respect to edges that change their congestion through time. We perform an experimental user study where we compare interactive and non-interactive interfaces, the complexity levels of the map structures (number of nodes and edges) and of the paths (number of nodes that need to be visited), and the effects of familiarity with the map. The results of our study indicate that an interactive interface is more beneficial than a non-interactive interface for more complex paths, while a non-interactive interface is more beneficial than an interactive interface for less complex paths. In detail, while the number of nodes and edges of the network had no effect on the performance, we observed that (not surprisingly) the more complex the path, the longer the processing time and the lower the correctness. We tested the familiarity with a test–retest design, where we organized a second session of tests, labeled T2, after the first session T1. We observed a familiarization effect in T2, that is, the participants' performance improved for the networks known from T1.*
Keywords: *spatio-temporal networks, route planning, user study, complexity, navigation*

[DOI: 10.2352/J.Percept.Imaging.2025.8.000401]

1. INTRODUCTION

Finding paths in spatial networks is one of the tasks that are more challenging using a matrix-based representation rather than using a node–link representation of the network [22, 35]. The choice of the right path may be complicated because the network conditions may change over time (e.g., the targets to reach may be "available" only in specific time windows throughout the day) or because network congestion may discourage the choice of some paths during the day. Still, deciding the sequence of places to visit and the paths that allow us to reach them is an ordinary task in everyday life (e.g., for tourists planning a route through a city). Choosing one route over another can have effects on the time spent commuting, which influences personal satisfaction, spent money, and pollution. Especially in highly crowded cities, congestions have a major impact on the travel time. Typically, users want to avoid congestion by accepting even large detours. Hence, investigating which visualizations work well in this context is essential. For example, Figure 1 shows two congestion conditions for the streets of a big city forecast by Google Maps. A user who has to perform different errands during the day would need to intersect the space and time information provided by the two maps. In addition, since users may have to select their optimal routes within a city over multiple days, familiarity could also play a significant role. From a more methodological perspective, if familiarization turns out to have a positive effect on the measured user performance, it can be assumed that in real-world applications, a training phase would be proportionally beneficial.

The domain of dynamic networks has been deeply studied (refer to [3, 9] for a taxonomy of models and solutions, and to [18] for an overview of surveys). In this context, several user studies consider descriptive and diagnostic analytic tasks involving paths in dynamic graphs; see, for example, [5–7, 39, 43].

The focus of this paper lies on the predictive and prescriptive visual analysis that is necessary for forecasting the interplay between space and time. This becomes particularly pertinent when traversing a path in a map through time. To our knowledge, this scenario, which we call *time-dependent route planning*, has not been explored yet. A high number of application contexts, instead, exhibit the interaction between time-dependent events and space-embedded networks. Hence, the findings of our investigation could be beneficial for the design of visual tools to support analytic tasks on dynamic transportation networks such as metro lines, railways, airline networks, container shipping networks, and so on. We remark that most of the routing and scheduling problems involving such networks are NP-hard problems [28]. Therefore, automatic computation is not always recommended, and human interaction via visual interfaces could be effective in finding sub-optimal solutions. We describe an experimental study aimed at determining strategies to assist users in managing predictive and prescriptive tasks within a spatio-temporal setting.

More precisely, we present to the users the network and its congestion in each time window. Then, we ask them to

Received Apr. 19, 2024; accepted for publication Dec. 1, 2024; published online Feb. 18, 2025. Associate Editor: Cindy Xiong Bearfield.
2575-8144/2025/8/000401/12/$00.00

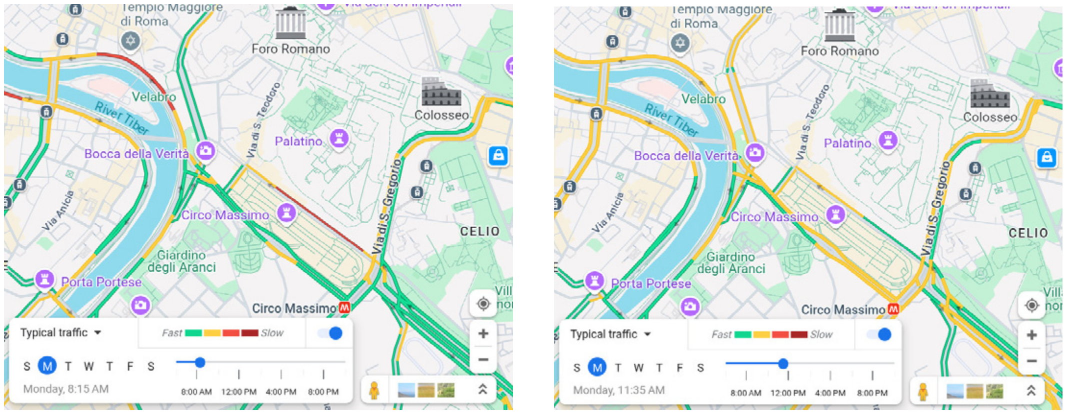


Figure 1. A typical application scenario: two congestion conditions for the traffic in Rome according to Google Maps (Map data ©2024 Google).

choose among three alternative paths the one with the least congestion at the moment the different roads along it have to be traversed. The difficulty for the user is to take into account both the spatial variables (the positions reached on the map) and the temporal variables (the congestion conditions varying over time).

We compare two different user interfaces: an interactive and a non-interactive version, where the former interface maintains the spatio-temporal nature of the problem while the latter maps it into a sequence of purely spatial subproblems. Furthermore, we ask the participants to repeat the experiment with a break in between (two test phases overall). We also use different complexity levels of the scenario: *map complexity* (number of edges and nodes of the network) and *path complexity* (number of target nodes in the paths). Our hypotheses are that for less complex scenarios, the non-interactive interface works better and that there is a familiarization effect in the second test phase. For a more formal definition of our research hypotheses and the experimental design, we refer to Sections 3 and 4, respectively.

The main contributions of this paper are as follows:

- the design of a user study of a composite (high-level) routing task for a complex scenario with spatio-temporal networks, which aims at
  - investigations of the effectiveness and user preference of interactive and non-interactive interfaces,
  - investigations of user familiarity achieved by two test phases;
- the statistical proof that simple scenarios are better supported by non-interactive interfaces while complex scenarios are better supported by interactive interfaces; and
- the statistical proof of familiarity effects.

2. RELATED WORK AND BACKGROUND

2.1 *Visual Analysis Paradigms*

The classification of the purposes of analysis into descriptive, diagnostic, predictive, and prescriptive was initially established within economic and the financial realms [20, 47]. Subsequently, it was adopted in biomedical research [26, 31] and visual analytics [2, 29]. Descriptive analytics answers the question "What happened?"; diagnostic analytics answers the question "Why did it happen?"; and predictive analytics focuses on "What is likely to happen?". Prescriptive analytics, instead, is much more ambitious and aims to drive actions so that something will (or will not) happen. In this paper, we especially focus on predictive and prescriptive visual analytics. In particular, we consider to what extent users are able to choose paths that will be traversed through time, avoiding events that intertwine and intersect with the traversals themselves. In order to do this, users need to conceptualize a path into multidimensional space where one axis is time (see Figure 2(b)) rather than simply coping with a path that evolves through time (as in Fig. 2(a)).

2.2 *User Evaluations of Network Visualizations*

To the best of our knowledge, no user studies have yet been conducted on predictive and prescriptive visual analysis of networks. In this section, we review the literature about descriptive and diagnostic user evaluations of network visualizations.

In a recent paper [18], Filipov et al. have conducted an overview of surveys to provide researchers and practitioners a "roadmap" elaborating the current research trends in the field of network visualization. They categorize recent surveys and task taxonomies published in the context of network visualization. Purchase et al. [36–38] investigate the impact of aesthetics on user performance and human understanding of node–link diagrams of graphs. Ghoniem et al. [22] describe a taxonomy of generic network-related tasks and perform an evaluation aimed at assessing the readability of two representations of networks: matrix-based representations and node–link diagrams. The study shows that matrix-based visualizations perform better than node–link diagrams on most tasks when graphs have greater than 20 vertices. Only path finding is consistently in favor of node–link diagrams throughout the evaluation. Keller et al. [27] extend the study to directed graphs and Holten et al. [25] study how the

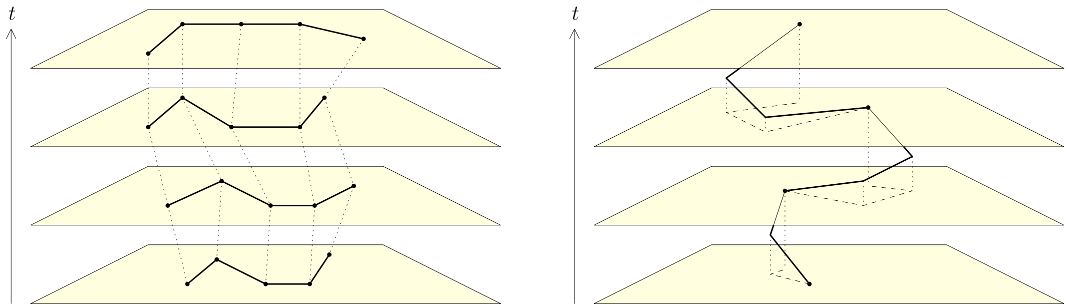


Figure 2. Two instances of paths with temporal information: (a) a descriptive or diagnostic visual analytic scenario involving a dynamic path; (b) a predictive or prescriptive visual analytic scenario involving a path traversed through time.

visual encoding of the direction of edges impacts human performance. More recently, the readability of network visualizations that combine the use of node–link diagrams with matrix-based representations has been investigated. More precisely, the study by Di Giacomo et al. [15] considers *hybrid* network visualizations, that is, node–link diagrams that contain sub-networks that are represented according to different drawing paradigms, such as adjacency matrices and chord diagrams [4, 24].

Didimo et al. [16] investigate the usability of overloaded orthogonal drawings (i.e., orthogonal drawings where "parallel" segments of two distinct edges can partially overlap) against classical orthogonal drawings, hierarchical drawings, and matrix-based representations for performing a collection of basic user tasks on directed graphs. Directed graphs are also the subject of the crowd-sourced user study by Abdelaal et al. [1], where node–link diagrams, adjacency matrices, and bipartite layouts are compared mainly focusing on overview tasks for large instances. See also the survey on empirical user evaluations of graph visualizations by Burch et al. [12].

Overall, there is a broad consensus among researchers that node–link representations of networks are the most effective when tasks requiring path analysis are considered. For this reason, in this paper we restrict our study to node–link representations.

2.3 *Dynamic Network Visualizations*

Surveys about the representations of dynamic networks can be found in the literature (see, e.g., [3, 9, 30]). The two main strategies for representing dynamic phenomena on a network are *time-to-space mapping* and *time-to-time mapping* [9]. The first strategy encodes the time dimension into some geometric object, that is, into a space dimension. This kind of representation may be very challenging for some application domains. A common technique, which we use in our experiments, is that of relying on small multiples, that is, replicating the representation for different discrete times. The second strategy maps the time dimension of the dataset into the time of the user, actually showing a dynamic view where the changes in the dataset through time are animated in simulated time. This kind of representation is also used in our experiments. Beck et al. [9] further classify as "hybrid" those systems that combine both the above described visualization strategies (closely connected) together. However, this happened only for a few systems where it was allowed by the application at hand.

Similar to our setting is the work described by Saraiya et al. [44], where a node–link diagram with static positions is used and only node attributes are time-varying (in our case, edge attributes also are). Comparing an animated slider solution to an approach with small time-series visualizations inside each node, they observe better performance by participants for the animated approach for tasks that involve only one or two points in time while the reverse happens for tasks involving more time steps. None of the tasks, though, asked for finding paths on the map.

Several user studies consider descriptive and diagnostic analytic tasks involving paths in dynamic graphs. Saffrey and Purchase attempted to identify when the shortest path between two nodes is shorter [43]. The focus, however, is on measuring the effects of the drawing stability (which seems to be of little help, if not a hindrance, for this task). Archambault et al. [5] investigated the difficulty of identifying the pair of nodes for which the shortest path always decreases. Small multiples are shown to be faster than animation but no significant difference is found in the error data. Archambault and Purchase [7] considered the task of recognizing a given path when the positions of the nodes change. With respect to small multiples, the animation is shown to produce answers closer to the correct one, and this is especially true for longer paths (five or seven nodes). Animation is preferable also in terms of response time.

Boyandin et al. [11] focus on the qualitative differences between the types of findings users make with animations and small multiples. They show that animation tends to reveal more findings on adjacent time steps while small multiples foster the discovery of patterns lasting over longer periods. Based on the above results, Beck et al. [9] conclude that small-multiple approaches seem to be preferable for tasks involving more than two time steps.

To our knowledge, the problems investigated in the literature about dynamic graphs do not address the case of predictive and prescriptive analytic tasks when some paths have to be traversed from the source to the destination forecasting hindrances during the traversal (refer to Fig. 2(b)), which is the subject of our study, but rather focus on paths that change through time (as illustrated in Fig. 2(a)). As already stated in the introduction, solving routing problems while anticipating events during route traversal is crucial to saving time and economic resources in several application scenarios.

2.4 *Geospatial Network Visualizations*

Geospatial network visualizations associate nodes and links with geographic locations either on Earth or other planets [45]. These visualizations are used to show, for example, trade and financial connections between countries and regions [8] or to display flight connections [42].

Schöttler et al. [45] present a systematic review of geospatial network visualization approaches by establishing a design space, which supports designers in building appropriate and effective visualizations for this type of network data. The proposed design space consists of the following dimensions: (i) geographical facet representation, (ii) network representation (for both nodes and edges), (iii) composition (how the topology and geography are combined in the visualization), and (iv) use of interaction. The geographical representation tackles how to encode geospatial information, which ranges from explicit (representations that use a cartographic map) to distorted (representations that use displacement of spatial positions according to some properties of the network), and to abstract (representations that use encodings not based on map projections) [14]. However, geospatial network visualization captures several open challenges, like handling co-located nodes, link density, and uncertainty in geospatial networks.

According to the survey by Schöttler et al. [45], only 7% of the papers in their collection captured an evaluation of geospatial network visualization techniques (mainly case studies and small quantitative user studies). This is also caused by a lack of structured task taxonomy for geospatial networks, and only basic tasks (such as identify and compare) are used. In our study, we aim to tackle this problem by investigating high-level tasks. Schöttler et al. [45] recommended utilizing their design space for obtaining empirical evidence according to categories.

2.5 *Familiarity in Network Visualization*

Familiarity effects have been studied in several research domains [19, 46, 49]. For the task of searching, for example, familiarity with targets or with distractors has been proved to be beneficial [32, 40, 41]. Frequently encountered patterns improve performances even when no recognition cognitive process is involved [49]. Familiarity is sometimes taken into account also in the network visualization research field [10]. In this domain, a concept related to familiarity is that of *mental map preservation* [17], which assumes that drawing similar graphs in a similar way enhances the user performance in a dynamic setting. However, the literature provides evidence that preserving the mental map is not always helpful when performing tasks on dynamic graphs [6], and for some specific tasks it can be detrimental [39, 43]. Archambault and Purchase [6] report the following: for local tasks, the number of considered nodes, edges, or sub-graphs needs to be high to observe a positive effect of mental map preservation; for global tasks that involve indistinguishable nodes and edges, no benefit is measurable; path finding, which is a global task involving distinguishable nodes and edges, is positively affected by mental map preservation when the number of elements to consider is high.

3. RESEARCH HYPOTHESES

In our experiment, the users have to cope with a composite task: projecting a path in time (see Fig. 2(b)) and anticipating what conditions will be met while traversing such a path. In order to answer these questions, the users would benefit from the conceptualization of a multidimensional space where one axis is time. The strategy of mapping time to space, adopted by non-interactive interfaces, corresponds to splitting the spatio-temporal problem into a collection of spatial-only sub-problems that are easier to answer singularly. The strategy of mapping time to time, adopted by interactive interfaces, maintains the temporal axis and leverages on the user multidimensional space conceptualization. The aim of our investigation is to understand which of the above two strategies is the most effective and under which conditions. Furthermore, we ask the participants to repeat the experiment with a break in between to evaluate the possible familiarization effect. In fact, in some application scenarios, users may have to select their optimal routes within a city over multiple days.

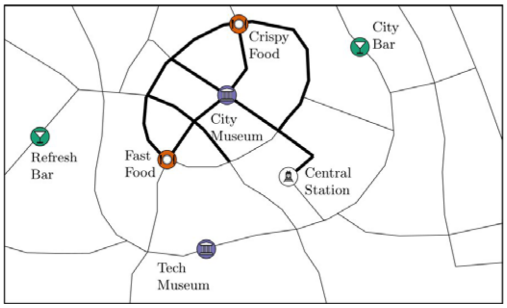
For our experiment, we formulate the following research hypotheses.

(H1) For less complex scenarios, the non-interactive interface leads to (a) higher correctness and (b) shorter processing time for correct answers.

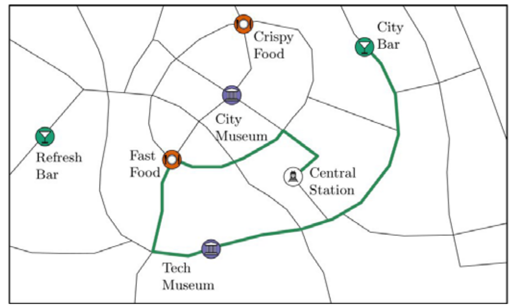
(H2) For more complex scenarios, the interactive interface leads to (a) higher correctness and (b) shorter processing time for correct answers.

(H3) In the second test phase, there is a familiarization effect. That is, for stimuli that were already shown in the first test phase, it shows (a) higher correctness and (b) shorter processing time for correct answers.

Recall that we define the complexity of a scenario by the map complexity (number of nodes and edges in our network) and path complexity (number of target nodes); see Section 4 for a formal definition. Hypotheses (H1) and (H2) are motivated by our intuition that for a complex scenario, the non-interactive interface is difficult for the participants since they need to manage the collection of sub-problems provided by the small multiples (e.g., they need to find the site locations over the different views) and then put together the answers to each sub-problem. On the other hand, we think that for a simple scenario, the overhead of the interactive interface leads to worse performance. Indeed, mapping time



(a)



(b)

Figure 3. (a) Stimuli of our experiment. Visualization of the congested roads for a commuting time window. Edges that are depicted with higher thickness indicate more traffic for the displayed time window. (b) The users are provided with three alternative options from which they should choose the optimal. Here, we display an exemplary one.

to time implicitly requires more time for the user to access the encoded information.

The effect of familiarization (hypothesis H3) is complex to predict since we aim for a time gap of a few days between the two test phases. We suspect that participants in the second test phase remember the stimuli that they already have seen before and perform better on these since solving the tasks requires a high degree of engagement and processing time.

4. EXPERIMENTAL DESIGN

In this section, we describe the design of our experiment. In particular, we discuss the application scenario we address, the visualization concept we adopt, the task of our experiment, the experimental factors we consider, the stimuli we generate, and the experimental setting.

4.1 *Application Scenario*

Inspired by the real-world situation where a tourist wants to plan a route to visit a city, the scenario we address is the following. The user is at a location corresponding to a point on the map and her purpose is to plan a route that allows her to visit a given sequence of site categories avoiding congested roads as much as possible, independent of their length. The scenario is dynamic, that is, all sites of the same category are available in a specific time window, and the congestion of each road can vary over time.

4.2 *Visualization Concept*

We represent a map as a node–link diagram of a network, whose edges correspond to roads and whose nodes correspond either to sites or to intersections between roads; refer to Figure 3 for an illustration. Each site–node is assigned a category (e.g., “restaurants,” “museums”) and is represented as a colored disk containing an icon; the category is encoded by the color and by the icon of the node to create an inclusive visualization for users with vision deficiencies such as color blindness. For the purpose of our study, each site belongs to a category and all nodes of the same category are equivalent to the user. For example, a restaurant is a site and there can be multiple nodes belonging to the “restaurants” category; for example, in Fig. 3, there are two restaurants called “Fast Food” and “Crispy Food.” Edges are represented as curves between nodes. A congested road is represented as a thicker curve. Since it is known that for exploratory tasks interactive interfaces may be beneficial [33], we consider both a non-interactive and an interactive interface. For the non-interactive interface, we adopt the small-multiple visualization paradigm, where representations of the map corresponding to different time windows are displayed next to each other (see Figure 5). Each of these representations shows the congestion situation of the roads in the time window where the user has to move from one site to another. In the interactive interface, the user can select the time window for which she wants to see the congestion situation (see Figure 6).

4.3 *Task*

We defined one task with the aim of reflecting a real-world situation where a tourist wants to plan a route to visit a city. The network is shown either with the non-interactive interface or with the interactive interface, and the task is the following: the participant is located at the central station and she wants to visit a given sequence of site categories by choosing the path that avoids congested roads as much as possible. Both the central station and the sites to visit are represented as nodes on the map as discussed above. For example, if the user wants to visit a restaurant, she has to choose the alternative that allows her to avoid congested roads; in Fig. 3(a), the most direct path that goes from the central station to the restaurant “Fast Food” is better than the one that goes to the restaurant “Crispy Food” since it contains a lower number of congested roads. To complete the task, the user is provided with three different paths (see, e.g., Fig. 3(b)) from which she can select the optimal one, that is, the one that avoids as many congested roads as possible, independent of their length.

4.4 *Experimental Factors*

Our experimental design consists of four fully crossed experimental factors; refer to Figure 4:

(a) (b)

Figure 4. Stimuli of different complexity levels: (a) less complex scenario (*NumAlt* = 2, *NumSites* = 3, *NetSize* = small); (b) more complex scenario (*NumAlt* = 3, *NumSites* = 5, *NetSize* = large).

(1) *Interactivity* of the interface (non-interactive versus interactive).
(2) *NumAlt*: Number of alternatives for each site (2 versus 3).
(3) *NumSites*: Number of sites to be visited (3 versus 5).
(4) *NetSize*: Size of the network in terms of the number of nodes and edges.
    (a) *Small*: corresponds to 15–30 nodes and 25–60 edges.
    (b) *Large*: corresponds to 5–20 nodes more and 15–50 edges more than those of small NetSize.

This leads to 16 different combinations of the experimental factors. We consider a scenario to be *less complex* than another scenario if at least one between *NumAlt*, *NumSites*, and *NetSize* is smaller while the other measures are not larger. Note that *NumSites* is also correlated to the complexity of the path of the solution options.

4.5 *Stimuli*

To avoid the effect of possible previous knowledge of cities by the users, the maps represent fictitious cities with sites having fictitious names. The maps were created manually by students with expertise in cartography. The students were instructed to create city structures that are real-world-like but that do not have any specialties such as bridges, tunnels, and one-way streets. Hence, the street network is an undirected planar graph. For every map, we generated versions for each combination of the experimental factors *NumAlt*, *NumSites*, and *NetSize*. For generating the congested edges per time window, we decided to focus on one or two congested areas rather than randomly distributing congested edges in the network; see Figs. 3(a), 5, and 6. Congested edges are depicted with a higher thickness to create an inclusive visualization for users with vision deficiencies such as color blindness. The three paths provided as answer options were generated manually such that the optimal route contains a congested road only if it is also contained in the other two options. We made it clear to the participants that avoidance of congested roads was the main and only task and that stumbling on a congested road would imply to be stuck for an unpredictable time (see task description at the top of Figs. 5 and 6). Although we tried to make the three options equally difficult for each map, there were some differences due to the city layouts and distribution of the sites. We point out that we designed the stimuli so that they are also usable for participants with visual deficiencies.

4.6 *Experimental Timeline*

We designed a within-participant experiment where the users were exposed to all the generated stimuli. The test was articulated in two test phases:

(T1) The steps of the first test phase were the following: (i) collection of demographic data about the participant; (ii) instructions for the execution of the test; (iii) training phase with some sample trials; (iv) presentation of 16 experimental trials (one for each combination of experimental factors) in random order; (v) collection of qualitative feedback (aesthetic appeal and impression, perceived complexity in performing the tasks, concentration level of the participant, strategy for solving the task).
(T2) In the second test phase, participants were exposed to 32 trials, consisting of the same stimuli from T1 plus 16 new stimuli. In this test phase, we also collected qualitative feedback as in T1.

The participants were asked to do both test phases with a break of 1–7 days in between. This allows them to be distracted from the main task and to regain full attention at T2. We estimated the processing time to be about 30 minutes for T1 and about 1 hour for T2.

4.7 *Participants*

Our target participant sample is a convenience sample, that is, participants without any specific knowledge. The relevant hypothesis for the power analysis, which determined the sample size, was hypothesis (H1a/b). As we base our analyses on multilevel data analysis using linear mixed models (LMMs), we employ R package simr [23] for power calculation. The main statistical model we use is based on a repeated measures design. The effect in question is about the additional fixed effects of *NumAlt*, *NumSites*, and *NetSize* on correctness, and it was set to $b1 = -5$, $b1 = -2$, and $b1 = -1$, respectively, which represent a mixture of small to

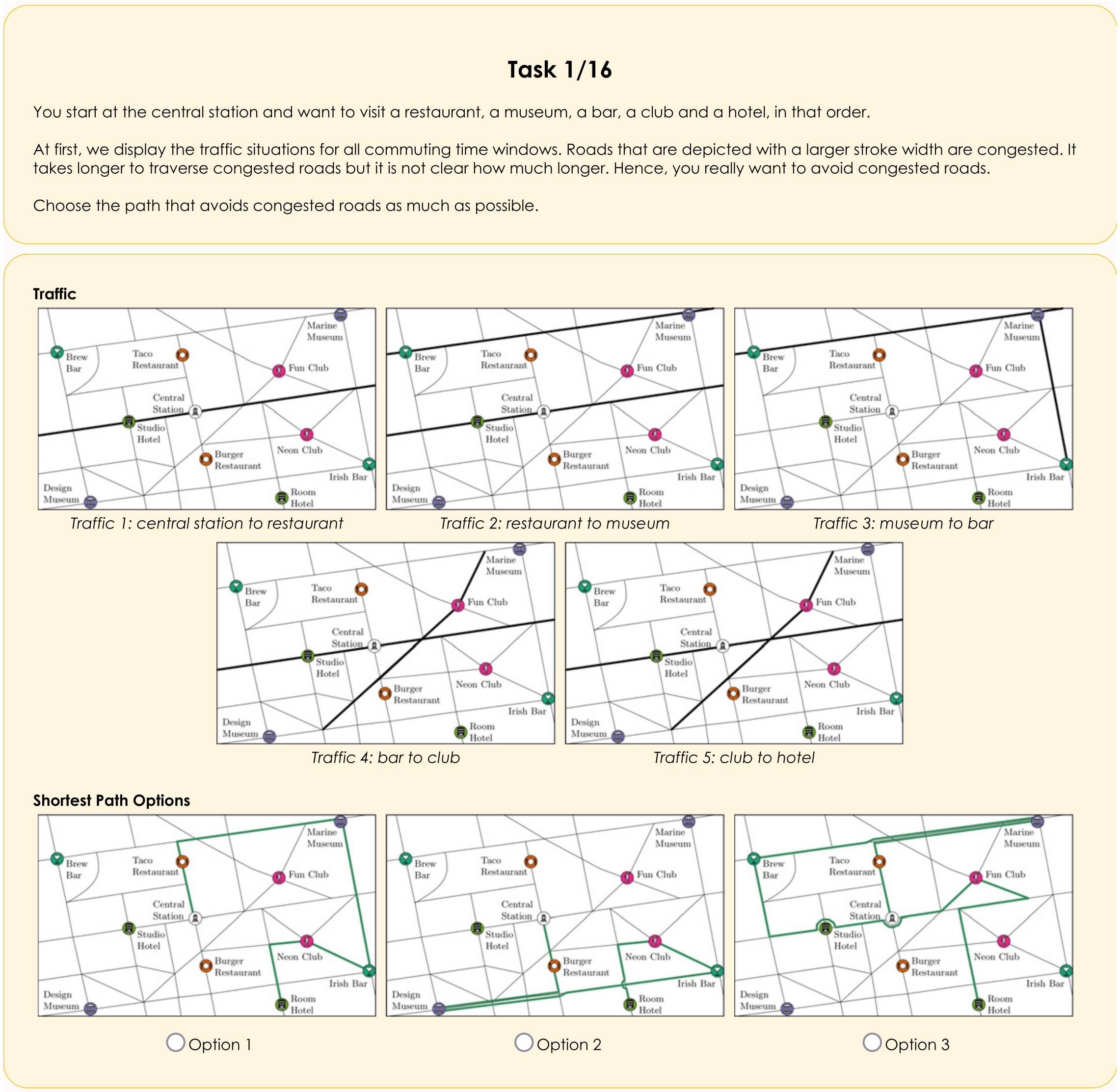


Figure 5. A screenshot of the user interface for a task with complexity level: *NumAlt* = 2, *NumSites* = 5, *NetSize* = large. At the top, we display the task description. In the middle, we display the traffic for each time window with the small-multiple representation (non-interactive). At the bottom, we display the three solution options.

medium effect sizes [34]. The b-parameters were oriented to typical effects found in spatial cognition and cognitive psychology as such. Such a priori assumptions always represent rough estimations; accordingly, the calculation of test power is always vague in the end but provides a range of needed participants to fulfill the pre-set criteria. To observe that this effect explains a significant amount of variance compared to the null model with $\alpha = 0.05$ and a satisfactory test power $1 - \beta$ of 0.80, we aim to collect data from $N = 53$ participants. As we expect a drop off of about 20% of the participants including persons who do not respond to the tasks adequately, we recruit $N = 64$ participants. We recruit participants without specific knowledge about routing in touristic scenarios via the online company Clickworker. Additionally, we recruit student participants and co-workers from the collaborating universities.

5. EXPERIMENT AND RESULTS

5.1 *Setup*

The participants of our study executed the test by accessing a web interface[1]; see Figs. 5 and 6. To reduce distractions and ensure proper visualization, the participants were asked to change to full-screen mode and ensure a reasonable screen size by providing a button that checks whether their screen size is at least 1200×800 pixels. We did not enforce this screen size or record the device type. The participants were allowed to take a break after each task. In order to test our study design, we performed a pilot study with three participants, which led only to minor changes in the instruction text. The study was online for 15 days, starting on May 15, 2023.

[1] https://optimierer.igg.uni-bonn.de/tourismrouting.

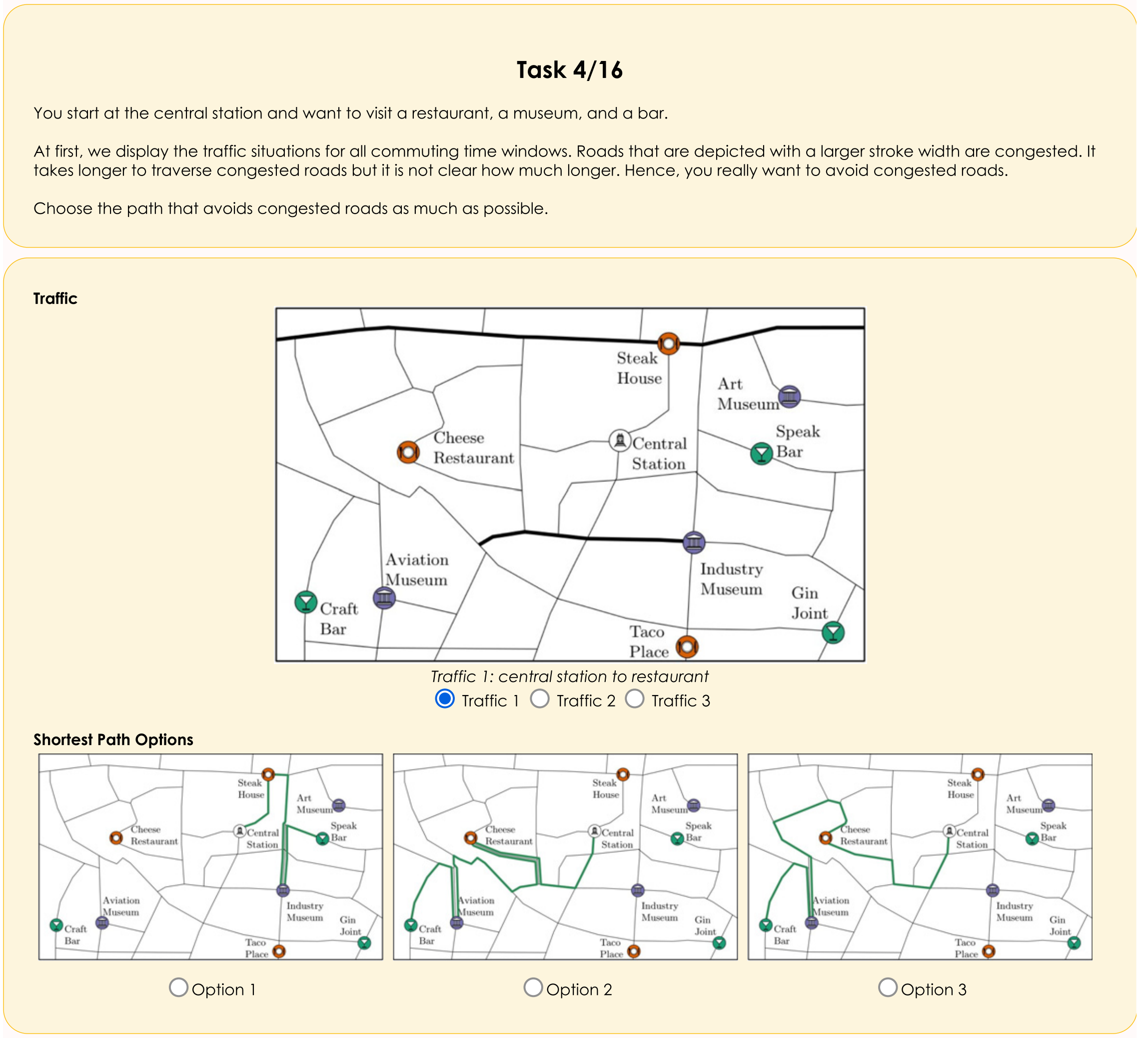


Figure 6. A screenshot of the user interface for a task with complexity level: *NumAlt* = 3, *NumSites* = 3, *NetSize* = large. At the top, we display the task description. In the middle, we display the traffic with the interactive interface. The user can interact with the interface using the buttons "Traffic 1",..., "Traffic 3" to look at the congestion for a different time window. At the bottom, we display the three solution options.

The final number of participants was $N = 78$ (28 female, 47 male, 2 other, 1 did not indicate their gender; mean age 32.4 years [19–58 years]); we used all datasets (one per person if participants took part only in test phase 1 or two if they took part in both test phases) that fulfilled the criterion that participants had responded to at least more than 50% of all trials (8/16 for T1 and 16/32 for T2) with at least 1/3 of correct responses. Most participants executed both parts in a sequence ($n_{1,2} = 49$); the rest executed only T1 ($n_1 = 29$). Thus, we exceeded the critical number of 53 participants taking part in test phase 1, which we calculated a priori to be able to reveal the assumed effects within test phase 1. The delay between the end of test phase 1 and the start of test phase 2 was 3 days and 7 hours on average.

5.2 *Measures*

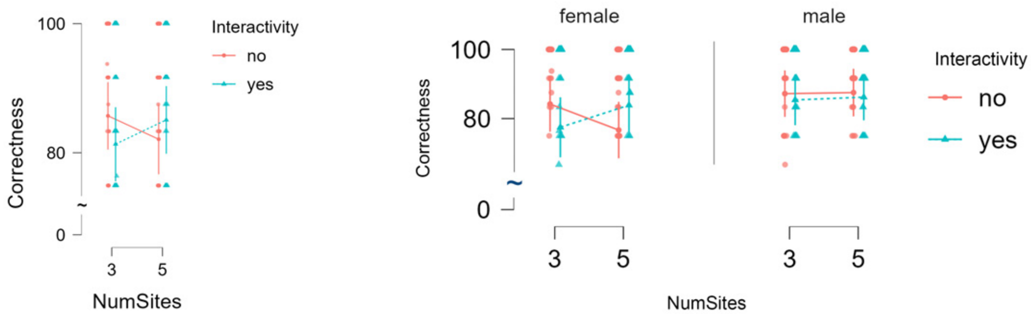
We evaluate the participant's performance using the following quantitative measures:

(a) correctness (ratio of the number of correct answers to the total number of questions);
(b) processing time for correct answers (time taken by a participant to answer—only in the case of correct answer); and
(c) familiarization effects by comparing the correctness in test phase 1 to test phase 2 and comparing the processing time in test phase 1 to test phase 2.

Lastly, we analyzed the quality of the study design by using two simple qualitative measures:

(d) the ease of use of the map tool (*How easy did you find the tasks?* [very difficult/difficult/neutral/easy/very easy]); and

(e) the degree of how much participants liked the interactive mode and the degree of how much participants liked the small-multiple mode (*How do you like the timeline interface (interactive version)?* [It is very bad/It is bad/It



(a) (b)

Figure 7. (a) Effects of *Interactivity* and *NumSites* on the correctness rate (research hypotheses (H1) and (H2)). (b) Differential effects of *Interactivity* and *NumSites* for female (left) and male participants (right) on the correctness rate (research hypotheses (H1) and (H2)).

is neutral/It is good/It is very good] and *How do you like the small-multiple interface (series of static images next to each other)?* [It is very bad/It is bad/It is neutral/It is good/It is very good]).

5.3 *Statistical Analysis of (H1) and (H2)*

We tested (a) and (b) on the different dependent variables utilizing separate LMMs. An LMM [21] is a statistical model that combines fixed effects (consistent, predictable factors) and random effects (variables that introduce random variation) to analyze data with complex, hierarchical structures, such as repeated measures or clustered data. The advantage of an LMM is its ability to handle dependencies within the data, leading to more accurate and reliable estimates in situations where traditional linear models may fall short, particularly in longitudinal or multilevel studies. In the present study, we used the participants as the random factor and the fixed factors *Gender* (only female versus male due to a too low number of other genders available in this dataset), *Interactivity*, *NumAlt*, *NumSites*, and *NetSize* as fixed effects. We included all main effects plus interactive effects of *Gender* and *Interactivity* with all other factors.

For the dependent variable (a) correctness, we obtained two-way interactions between *Interactivity* and *NetSize*, $p < 0.05$, and between *Interactivity* and *NumSites*, $p < 0.001$ (see Figure 7(a)), and a three-way interaction between *Gender*, *Interactivity*, and *NumSites*, $p < 0.05$ (see Fig. 7(b)).

To reveal further possible effects, we conducted more detailed analyses with regard to educational background, age, and gender. Although we did not find any effects regarding the educational background or the age, there is an interesting effect regarding the gender. Whereas female participants benefited from the interactive mode when five instead of three *NumSites* were used, we did not obtain any differential effect for male participants. We can only speculate at this point why female participants benefited from the interactive mode. Some gender-specific preferences and usability aspects have been researched so far, and such research is important for future considerations about gender-adequate interfaces, for instance. However a uniform picture of results is missing even though now quite complex ways of addressing such questions are available as recent research shows (see, e.g., Xu et al. [48]). But this would go beyond the scope of the present research study.

For the dependent variable (b) processing time of correct responses, we obtained only the main effect of *NumSites*: the more the number of sites available, the slower the participants in finding a solution ($p < 0.0001$).

We can conclude that our research hypotheses (H1) and (H2) are partly confirmed.

5.4 *Statistical Analysis of (H3)*

Regarding the sustainability of correctly processing the maps at test phase 1 transferred to test phase 2, we executed two separate LMMs, again with participants as the random factor and now with *Familiarity at T2* as the fixed effect (whether a map had been already familiarized through processing at T1). We again employed the two main quantitative measures, correctness and processing time of correct responses, as the dependent measures. We obtained clear effects of familiarization for both measures: correctness rate raised from 77.98% to 86.25%, $p < 0.0001$, and the processing times of correct responses dropped from 52.5 s to 39.7 s, $p < 0.001$. See Figure 8 for details.

We can conclude that our research hypothesis (H3) is confirmed.

5.5 *Statistical Analysis: More Findings*

For analyzing measures (d) ease of use and (e) degree of liking of the tools, we asked three questions about (1) the ease of use of the map tool, (2) the degree of how much participants liked the interactive mode, and (3) the degree of how much participants liked the small-multiple mode.

First, we evaluated these dependent variables by means of separate LMMs with the fixed factor *test phase* (T1 versus T2) and with participants as random factors. We obtained medium–high evaluations (between 2.93 and 3.72). Hence, we presume that the task's difficulty levels are appropriate and both interfaces are suitable for solving the task.

Second, we can test with these values whether the participants' opinions changed from T1 to T2. For example, one can suppose that the tasks are perceived easier in T2

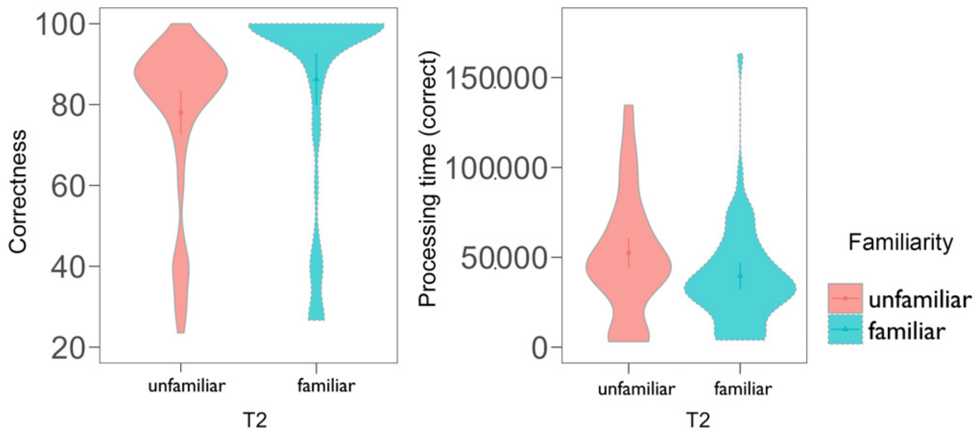


Figure 8. Effects of familiarization for the measures correctness (in %, left) and processing time (in ms, right)) of correct responses (research hypothesis (H3)).

due to more training or that participants like the interactive interface more in T2 since it is not new anymore. The test showed that there is no significant change for any variable between T1 and T2, $p > 0.35$. We presume that maybe more practice with the task and interface is needed to change the participants' opinions.

Third, we evaluate the selective liking of *interactive* ($M = 3.81$) versus *small-multiple* ($M = 3.60$) visualization. We conducted a paired Student t-test resulting in a non-significant difference between both variables ($t(84) = 1.56$, $p = 0.122$, Cohen's $d = 0.169$, n.s.); thus, there was no clear preference for either of these visualization modes. A subsequent correlation analysis between both variables (Pearson's $r = 0.008$, $p = 0.943$, n.s.) further showed that people who were in favor of one mode were not clearly devaluating the other mode, indicating relatively undecided preferences. Here, the participants' answers coincide with the evaluation of (a) correctness and (b) processing time where also we did not find one superior interface for all complexity levels.

5.6 *Additional Comments by the Study Participants*

In the following, we report findings that are unrelated to our explicit research hypotheses. We allowed the users to give comments on the user study in a free-text field. Two users suggested merging the traffic of all time windows into one visualization, for example, with different colors. We agree that one static visualization has its advantages but see a clear limitation concerning the number of time windows. Two users mentioned that the interactive interface is more enjoyable while the static is better considering the performance. Another user confirmed our research hypothesis that the static interface is more appropriate for fewer time windows and the interactive interface is better suited for more time windows.

6. CONCLUSIONS AND LESSONS LEARNED

In this paper, we investigated a composite user task that combines predictive and prescriptive visual analytics on spatio-temporal networks to support prediction in the interplay of time and space. Our experimental user study compared and evaluated an interactive and a non-interactive interface for this user task.

Statistical analysis shows that the non-interactive interface leads to higher correctness for simple scenarios (hypothesis H1) and the interactive interface leads to higher correctness for complex scenarios (hypothesis H2). If considering the gender factor, the described phenomenon is observed only for female participants and not for male participants. As a guideline for interface designers, our findings suggest that when the task implies forecasting the interaction of space and time and the scenario is not very simple, an interactive interface may be recommended. Furthermore, we investigated familiarization effects. In the second test phase, the participants were able to solve tasks for maps that they already saw in the first test phase more accurately while also being faster (hypothesis H3). This suggests, for example, that for the kind of task addressed in this study, a training phase on the map is expected to improve user performance.

From performing this study, we emphasize the following lessons learned:

- Crowdsourcing platforms, like Clickworker, facilitated a higher number of participants. Particularly the combination of employing participants from platforms like Clickworker and invited individuals supported the diversity of our participants but could also lead to biased results due to uneven samples in between-participant designs (e.g., characterized by different expertise, familiarity, or motivation levels).
- The careful, clear, and simple interface design of our experiment assisted in conducting our user study even on small and low-resolution displays. However, sometimes the participants reported difficulties seeing all alternative options at once—in future research, the screen size should be recorded and also determined whether a screen can be approached interactively or merely passively.

- The non-interactive interface led to higher correctness for simple scenarios and the interactive interface led to higher correctness for complex scenarios.
- Familiarization had an impact on the participants' performance. Therefore, it can be assumed that people who are familiar with spatial structures (e.g., city maps) can process and grasp a significantly greater level of complexity more quickly.

Besides tourist planning, which is the focus of this study, we believe that this work advances the visualization field, fostering its application to other dynamic transportation networks, such as metro lines, railways, airline networks, container shipping networks, and so on, by showing that interactive interfaces are recommended whenever the user is challenged by non-trivial instances that demand forecasting the interaction of space and time. Furthermore, the results on familiarization suggest that the ability of the users to take advantage of interactive interfaces may sum up through time.

We would also like to point out the limitations of our user study. Limitations are always present for user studies as you always have to find a balance in the design, for example, between the number of experimental factors and the duration of the user study. Experimental factors that we did not evaluate but that could have had an influence on our results are, for example, the type of device used, the screen size, how the participant gives the answer (e.g., drawing the solution herself instead of using the provided options), and different levels of traffic. We also want to make clear that the embedment of such routines and the concrete application–for what sake it will be used—will make a difference, so further contextualized evaluations will be required to check the practicability and usability of different parameters in the field. We see investigations of these experimental factors, especially with naturalistic or natural settings, as interesting follow-up work.

Besides the above-discussed future work based on the mentioned limitations, we came across other interesting related problems.

- The assessment and analysis of further qualitative measures (e.g., impression, complexity, difficulty, and usefulness) could give more insight into user perception.
- Another hypothesis that we suggest testing is that visualizations that are more aesthetically appealing lead to higher familiarization effects.
- Investigating the user's attention behavior through eye tracking might give insights to improve the interfaces.
- Considering even more complex composite tasks such as letting the participants draw their preferred route instead of giving options would be interesting.
- Overall, we deem that it is important to investigate how to design maps that resemble better the psychologically plausible qualities of cognitive maps as already stated by Carbon [13].

ACKNOWLEDGMENT

This work was initiated at the Schloss Dagstuhl (Leibniz-Zentrum für Informatik) Seminar 23051 "Perception in Network Visualization," and is partially funded by (i) German Research Foundation under Germany's Excellence Strategy, EXC-2070-390732324 —PhenoRob; (ii) MUR PRIN Proj. 2022TS4Y3N— "EXPAND: scalable algorithms for EXPloratory Analyses of heterogeneous and dynamic Networked Data"; (iii) MUR PON Proj. ARS01_00540—"RASTA: Realtà Aumentata e Story-Telling Automatizzato per la valorizzazione di Beni Culturali ed Itinerari"; and (iv) FWF grant ArtVis (10.55776/P35767).

REFERENCES

[1] M. Abdelaal, N. D. Schiele, K. Angerbauer, K. Kurzhals, M. Sedlmair, and D. Weiskopf, "Comparative evaluation of bipartite, node-link, and matrix-based network representations," IEEE Trans. Vis. Comput. Graph. **29**, 896–906 (2023).

[2] Z. Ahmad, S. Yaacob, R. Ibrahim, F. Wan, and F. Wan, "The review for visual analytics methodology," *2022 Int'l. Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (IEEE Computer Society, Ankara, Turkey, 2022), pp. 1–10.

[3] J. Ahn, C. Plaisant, and B. Shneiderman, "A task taxonomy for network evolution analysis," IEEE Trans. Vis. Comput. Graph. **20**, 365–376 (2014).

[4] L. Angori, W. Didimo, F. Montecchiani, D. Pagliuca, and A. Tappini, "Hybrid graph visualizations with chordlink: algorithms, experiments, and applications," IEEE Trans. Vis. Comput. Graph. **28**, 1288–1300 (2022).

[5] D. Archambault, H. Purchase, and B. Pinaud, "Animation, small multiples, and the effect of mental map preservation in dynamic graphs," IEEE Trans. Vis. Comput. Graph. **17**, 539–552 (2011).

[6] D. Archambault and H. C. Purchase, "The 'map' in the mental map: Experimental results in dynamic graph drawing," Int. J. Hum. Comput. Stud. **71**, 1044–1055 (2013).

[7] D. Archambault and H. C. Purchase, "Can animation support the visualisation of dynamic graphs?," Inf. Sci. **330**, 495–509 (2016).

[8] A. Arleo, C. Tsigkanos, R. A. Leite, S. Dustdar, S. Miksch, and J. Sorger, "Visual exploration of financial data with incremental domain knowledge," Comput. Graph. Forum **41**, 101–116 (2023.).

[9] F. Beck, M. Burch, S. Diehl, and D. Weiskopf, "A taxonomy and survey of dynamic graph visualization," Comput. Graph. Forum **36**, 133–159 (2017).

[10] R. Blue, C. Dunne, A. Fuchs, K. King, and A. Schulman, "Visualizing real-time network resource usage," in *Visualization for Computer Security*, edited by J. R. Goodall, G. Conti, and K.-L. Ma (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 119–135.

[11] I. Boyandin, E. Bertini, and D. Lalanne, "A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples," Comput. Graph. Forum **31**, 1005–1014 (2012).

[12] M. Burch, W. Huang, M. Wakefield, H. C. Purchase, D. Weiskopf, and J. Hua, "The state of the art in empirical user evaluation of graph visualizations," IEEE Access **9**, 4173–4198 (2021).

[13] C.-C. Carbon, "Cognitive continental drift: how attitudes can change the overall pattern of cognitive distances," Environ. Plan. A **42**, 715–728 (2010).

[14] C.-C. Carbon and V. M. Hesslinger, "Attitudes and cognitive distances: On the non-unitary and flexible nature of cognitive maps," Adv. Cogn. Psychol. **9**, 121–129 (2013).

[15] E. Di Giacomo, W. Didimo, G. Liotta, F. Montecchiani, and A. Tappini, "Comparative study and evaluation of hybrid visualizations of graphs," IEEE Trans. Vis. Comput. Graph. **30**, 3503–3515 (2023).

[16] W. Didimo, F. Montecchiani, E. Pallas, and I. G. Tollis, "How to visualize directed graphs: a user study," *IISA 2014, The 5th Int'l. Conf. on Information, Intelligence, Systems and Applications* (IEEE Computer Society, Chania, Crete, Greece, 2014), pp. 152–157.

[17] P. Eades, W. Lai, K. Misue, and K. Sugiyama, "Preserving the mental map of a diagram," *Proc. Compugraphics '91* (Sesimbra, Portugal, 1991), pp. 24–33.
[18] V. Filipov, A. Arleo, and S. Miksch, "Are we there yet? A roadmap of network visualization from surveys to task taxonomies," Comput. Graph. Forum **42**, e14794 (2023).
[19] J. Flowers and D. Lohr, "How does familiarity affect visual search for letter strings?," Perception Psychophysics **37**, 557–567 (1985).
[20] Gartner Inc., "What is data and analytics?" available at https://www.gartner.com/en/topics/data-and-analytics. Access date March 16, 2024. [Online]. Available: https://www.gartner.com/en/topics/data-and-analytics.
[21] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, 2006).
[22] M. Ghoniem, J.-D. Fekete, and P. Castagliola, "A comparison of the readability of graphs using node-link and matrix-based representations," *IEEE Symposium on Information Visualization* (IEEE, Piscataway, NJ, 2004), pp. 17–24.
[23] P. Green and C. J. MacLeod, "SIMR: an R package for power analysis of generalized linear mixed models by simulation," Methods Ecology Evol. **7**, 493–498 (2016).
[24] N. Henry, J.-D. Fekete, and M. Mcguffin, "NodeTrix: hybrid representation for analyzing social networks," IEEE Trans. Vis. Comput. Graph. **13**, 1302–1309 (2007).
[25] D. Holten, P. Isenberg, J. J. van Wijk, and and J. Fekete, "An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs," in *IEEE Pacific Visualization Symposium, PacificVis 2011, Hong Kong, China, 1-4 March, 2011*, edited by G. D. Battista, J. Fekete, and H. Qu (IEEE, Piscataway, NJ, 2011), pp. 195–202.
[26] C. Hung, M. Tu, T. Chien, C. Lin, J. Chow, and W. Chou, "The model of descriptive, diagnostic, predictive, and prescriptive analytics on 100 top-cited articles of nasopharyngeal carcinoma from 2013 to 2022: Bibliometric analysis," Medicine **102**, e32824 (2023) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9907932/.
[27] R. Keller, C. M. Eckert, and P. J. Clarkson, "Matrices or node-link diagrams: which visual representation is better for visualising connectivity models?," Inf. Vis. **5**, 62–76 (2006).
[28] J. K. Lenstra and A. H. G. R. Kan, "Complexity of vehicle routing and scheduling problems," Networks **11**, 221–227 (1981).
[29] Y. Lu, R. Garcia, B. Hansen, M. Gleicher, and R. Maciejewski, "The state-of-the-art in predictive visual analytics," Comput. Graph. Forum **36**, 539–562 (2017).
[30] O. Michail, "An introduction to temporal graphs: An algorithmic perspective," Internet Math. **12**, 239–280 (2016).
[31] C. E. Morr and H. Ali-Hassan, *Analytics in Healthcare: A Practical Introduction*, Springer Briefs in Health Care Management and Economics (Springer, Cham, Switzerland, 2019).
[32] R. Mruczek and D. Sheinberg, "Distractor familiarity leads to more efficient visual search for complex stimuli," Perception Psychophysics **67**, 1016–1031 (2005).
[33] T. Munzner, *Visualization Analysis and Design*, A K Peters/CRC Press Visualization Series (CRC Press, Boca Raton, Florida, United States, 2015).
[34] P. Nieminen, "Application of standardized regression coefficient in meta-analysis," BioMedInformatics **2**, 434–458 (2022).
[35] M. Okoe, R. Jianu, and S. G. Kobourov, "Node-link or adjacency matrices: Old question, new insights," IEEE Trans. Vis. Comput. Graph. **25**, 2940–2952 (2019).
[36] H. C. Purchase, "Which aesthetic has the greatest effect on human understanding?," in *Graph Drawing, 5th Int'l. Symposium, GD '97, Proc.*, edited by G. D. Battista, Lecture Notes in Computer Science (Springer, Berlin, Heidelberg, 1997), Vol. 1353, pp. 248–261.
[37] H. C. Purchase, R. F. Cohen, and M. I. James, "Validating graph drawing aesthetics," in *Graph Drawing, Symposium on Graph Drawing, GD '95, Proc.*, edited by F. Brandenburg, Lecture Notes in Computer Science (Springer, Berlin, Heidelberg, 1995), Vol. 1027, pp. 435–446.
[38] H. C. Purchase, R. F. Cohen, and M. I. James, "An experimental study of the basis for graph drawing algorithms," ACM J. Exp. Algorithmics **2**, 4 (1997).
[39] H. C. Purchase and A. Samra, "Extremes are better: Investigating mental map preservation in dynamic graphs," in *Diagrammatic Representation and Inference, 5th Int'l. Conf., Diagrams 2008, Proc.*, edited by G. Stapleton, J. Howse, and J. Lee, Lecture Notes in Computer Science (Springer, Herrsching, Germany, 2008), Vol. 5223, pp. 60–73.
[40] X. Qin, W. Koutstaal, and S. Engel, "The hard-won benefits of familiarity in visual search: Naturally familiar brand logos are found faster," Attention Perception Psychophysics **76**, 914–930 (2014).
[41] R. Rauschenberger and H. Chu, "The effects of stimulus rotation and familiarity in visual search," Perception Psychophysics **68**, 770–775 (2006).
[42] P. Rodgers, "Chapter 7 - graph drawing techniques for geographic visualization," *Exploring Geovisualization* (Elsevier, Oxford, 2005), pp. 143–158.
[43] P. Saffrey and H. C. Purchase, "The 'mental map' versus 'static aesthetic' compromise in dynamic graphs: A user study," in *User Interfaces 2008, Ninth Australasian User Interface Conf., AUIC 2008*, edited by B. Plimmer and G. Weber, CRPIT (Australian Computer Society, 2008), Vol. 76, pp. 85–93, [Online]. Available: http://crpit.scem.westernsydney.edu.au/abstracts/CRPITV76Saffrey.html.
[44] P. Saraiya, P. Lee, and C. North, "Visualization of graphs with associated timeseries data," *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005* (IEEE, Piscataway, NJ, 2005), pp. 225–232.
[45] S. Schöttler, Y. Yang, H. Pfister, and B. Bach, "Visualizing and interacting with geospatial networks: a survey and design space," Comput. Graph. Forum **40**, 5–33 (2021).
[46] F. Tong and K. Nakayama, "Robust representations for faces: evidence from visual search," J. Exp. Psychol.: Human Perception and Perform. **25**, 1016–1035 (1999).
[47] C. Wolvius, "Predict the future, understand the past: the four types of data analysis," July 2020, available at: https://datajourney.akvo.org/blog/the-four-types-of-data-analysis. Access date March 16, 2024.
[48] W. Xu, N. Zhang, and M. M. Wang, "The impact of interaction on continuous use in online learning platforms: a metaverse perspective," *Internet Research* (Emerald Publishing Limited, 2023), Vol. 34, pp. 79–106.
[49] P. Yang and D. Beck, "Familiarity influences visual detection in a task that does not require explicit recognition," Attention, Perception, and Psychophysics **85**, 1127–1149 (2023).