

Science map metaphors: a comparison of network versus hexmap-based visualizations

Katy Börner^{1,2,3} · Adam H. Simpson¹ · Andreas Bueckle¹ · Robert L. Goldstone^{3,4}

Received: 24 October 2017
© Akadémiai Kiadó, Budapest, Hungary 2018

Abstract Most maps of science use a network layout; few use a landscape metaphor. Human users are trained in reading geospatial maps, yet most have a hard time reading even simple networks. Prior work using general networks has shown that map-based visualizations increase recall accuracy of data. This paper reports the result of a comparison of two comparable renderings of the UCSD map of science that are: the original network layout and a novel hexmap that uses a landscape metaphor to layout the 554 subdisciplines grouped into 13 color-coded disciplines of science. Overlaid are HITS metrics that show the impact and transformativeness of different scientific subdisciplines. Both maps support the same interactivity, including search, filter, zoom, panning, and details on demand. Users performed memorization, search, and retrieval tasks using both maps. Results did not show any significant differences in how the two maps were remembered or used by participants. We conclude with a discussion of results and planned future work.

Keywords Mapping science · Interactive data visualizations · User studies

✉ Katy Börner
katy@indiana.edu

¹ Cyberinfrastructure for Network Science Center, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA

² Indiana University Network Science Institute, Bloomington, IN, USA

³ Cognitive Science Program, Indiana University, Bloomington, IN, USA

⁴ Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Introduction

For centuries, cartographic maps have guided human exploration. Recent advances in data, algorithms, and computing infrastructures make it possible to map humankind's collective scholarly knowledge and technology expertise in science maps. Science maps render different scientific disciplines as "continents". Zooming into a discipline (e.g., mathematics, physics, or medicine) makes subdisciplines visible. Basemaps of science are commonly generated by analyzing citation links between millions of publications. Data overlays (e.g., showing all publications by one scholar, institution, or country) are computed by positioning records based on topical similarity.

Science maps are widely used to compare expertise profiles of different scholars, institutions, or countries; to understand career trajectories; to communicate emerging areas; or to visualize the diffusion of ideas or technologies. Prior work on the validity, utility, and legibility of science maps has compared different data sources for basemap generation, different data analysis and clustering methods, and different layout algorithms. Previous research has also compared the legibility of network versus map-based layout of general networks (see details in "[Prior work](#)" section). This paper documents the first comparison of two comparable maps of science reference systems plus interactive data overlays. User studies with 47 participants were run to evaluate how naïve users read, memorize, and use these maps.

The paper is organized as follows: "[Prior work](#)" section reviews prior work. "[Science basemap construction, data overlays, interactivity](#)" section details the construction of both basemaps, the generation of data overlays, and interactive functionality. "[Science map comparison](#)" section describes the experimental setup, data analysis, and results of the user study. "[Discussion and outlook](#)" section discusses the implications of the results and planned work.

Prior work

Just like early maps of the world were incomplete and partially erroneous, current maps of science are imperfect. The reasons are manifold: data is richer and more complete for certain disciplines of science (e.g., physics and medicine) and less complete for other areas of scholarly pursuit (e.g., arts and humanities that advance via books, physical artifacts, or performances); the properties of algorithms used to preprocess data (e.g., to unify author, geolocation, or journal names), extract networks (via citations or topical similarity), cluster data (partition or hierarchy), and lay out data (as a network or using a map metaphor) may only be partially understood. However, the increasing demand for and usage of science maps necessitates a detailed examination of what data, algorithms, workflows, visual metaphors and interactivity is best for a specific user group and task.

Early, pioneering work by Garfield (1955) focused on the compilation of citation indexes in support of the large-scale mining and mapping of science in support of students choosing a career in science (1974), mapping biochemistry and molecular biology for the years 1979–1980 (1981), mapping nutrition research (1987), epidemiology (1988), and biotechnology as well as molecular genetics (1984).

Hundreds of maps now exist, generated for different purposes, see review by Börner et al. (2003), the special *PNAS* issue on *Mapping Knowledge Domains* edited by Shiffrin and Börner (2004), and examples and detailed descriptions in Chaomei (2003), Börner's *Atlas* series

(2010, 2015), and the *Places & Spaces: Mapping Science* exhibit (2005–2017) featuring more than a hundred printed and several interactive large-scale maps (<http://scimaps.org>).

However, the large and diverse number of different methods applied to generate basemaps of science, to compute data overlays, and to support interactive exploration is rather confusing to the many users of science maps. Scientific studies that compare methods, algorithms, and visual metaphors are needed to identify what basemap, data overlay and interactivity are best for what user group and insight need.

Recent studies have examined the utility of different data sources for basemap generation (Boyack et al. 2011) and the consensus layout of different map making approaches (Boyack and Klavans 2009). Studies have also compared clustering algorithms for identifying fields in science (Boyack et al. 2005, 2009), some also using text-based similarity approaches (Boyack et al. 2011) to understand if citation linkages are needed to compute locally and globally accurate maps. Other work examined the scalability of different clustering algorithms when generating comprehensive maps of science (Emmons et al. 2016). Work by Skupin et al. (2013) evaluated the utility and legibility of self-organizing maps for mapping and understanding the structure of science. Prior work on the general legibility of network versus map-based visualizations of data showed that map-based visualizations do increase recall accuracy of data (Saket et al. 2014, 2015). Other work compared animations with static visualizations of temporal graph layouts (Farrugia and Quigley 2011). Work by Ghani and Elmqvist (2011) shows that adding static spatial features (e.g., background features, node size and color) to network visualizations helps with navigation and memorization.

To our knowledge, our work is the first to report the results of a comparison of two comparable interactive maps of science. This is partially due to the fact that the design of any interactive online visualization is rather time consuming and expensive. However, in order to truly understand the impact of visual metaphors on the utility and legibility of science maps, complete implementations are necessary.

Science basemap construction, data overlays, interactivity

The work we present here was conducted in close collaboration with economists within an NIH-funded project that aims to define, validate, and communicate existing and novel measures of scientific impact and transformativeness. Two interactive science map interfaces were implemented for the project to support the exploration and comparison of different metrics that aim at identifying high-impact and transformative science (HITS) (Staudt et al. 2016). The interfaces support the interactive visual display and comparison of up to four HITS metrics (normalized to values between 0 and 1) applied to any consecutive number of years between 1959 and 2016.

This section details the design of the two science basemaps, the data overlays and interactivity common to both maps, and the administrative interface that supports the upload and management of metrics.

UCSD basemap

The design of the UCSD map and classification system is detailed in (Börner et al. 2012). The map organizes publications in journals indexed by Thomson Reuters' Web of Science and Elsevier's Scopus into 554 subdisciplines. The subdisciplines are aggregated into 13 disciplines which are color-coded and labeled—e.g., *Surgery* and *Intensive Care* are aggregated

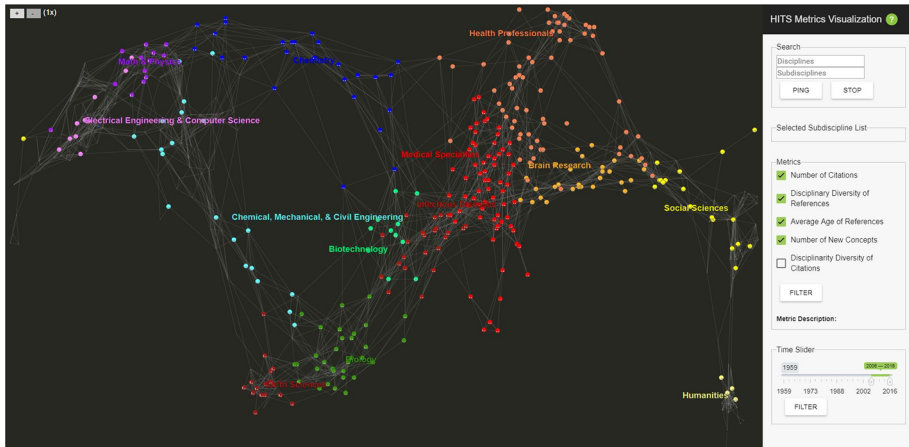


Fig. 1 Network visualization of the UCSD basemap. The panel on the right supports the selection of a time frame (here 2006–2016) and different metrics (four are selected). Hovering over a metric, e.g., *Disciplinary Diversity of References*, brings up an explanation, here “Herfindahl index of an article based on the fields of its references”. Explore the interactive version at <http://cns.iu.edu/econ-p/ucsdmap>

into *Medical Specialties* while *Semiotics* and *Linguistics* are aggregated into *Humanities*. Each of the 554 sub-disciplines has a set of journals and keywords associated with it. Links in the original ‘network’ rendering of the map (see Fig. 1) indicate journal co-citations (Fig. 2).

Hexmap basemap

Using the spatial positions of the 554 subdisciplines and 13 disciplines of science from the UCSD map and classification system discussed in the previous subsection, a hexmap was implemented as an alternative basemap for the comparison of HITS metrics. The hexmap layout started with the original node positions shown in Fig. 1; then, node positions were altered to give each node the same space. That is, areas with many nodes expanded in size; those with few nodes condensed. Nodes were rendered as hexagons and the layout was further modified to arrive at the regular honeycomb pattern shown in Fig. 3.

Data overlays and map interactivity

Both visual interfaces employ the widely used UCSD Map of Science (Börner et al. 2012) as a basemap and both support almost identical data overlays and interactivity.¹

The basemap can be explored by hovering over a discipline name, which highlights all of its subdisciplines (this only works for the hexmap); hovering over a subdiscipline

¹ The original UCSD network map was designed for print. The hexmap leverages the interactive functionality of the web. For print uses, the UCSD network map excels because all data is present at a single glance. The hexmap is more exploratory, inviting the users to interact with the initially visible data to explore details. A good example are the few *Health Professional* nodes that intrude on the *Medical Specialties* continent. On the hexmap, we can hover over the *Medical Specialties* nodes to see their names, see what connections to *Medical Specialties* and *Health Professionals* exists, and even click to find publications to determine why it is spatially similar to *Medical Specialties*. While the UCSD network map doesn’t prevent this, the hexmap was designed to promote this functionality.

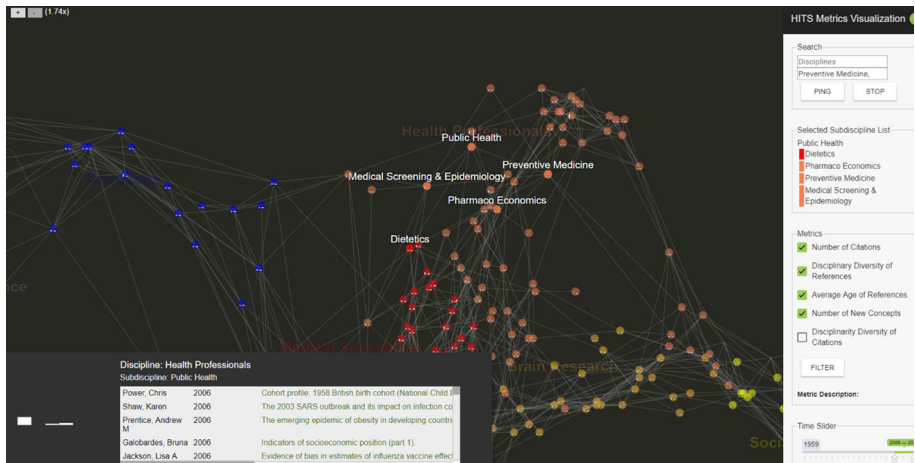


Fig. 2 UCSD basemap zoomed into the *Health Professionals* discipline. Hovering over the *Public Health* subdiscipline highlights and brings up labels for all subdisciplines linked to it. Clicking on it brings up a window at the lower left that lists publications with the highest combined metrics score of the four HITS metrics selected in the panel on right

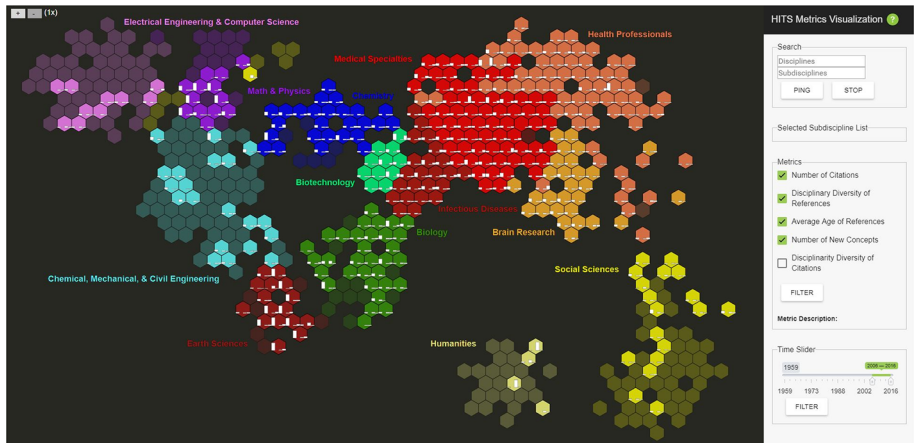


Fig. 3 Hexmap visualization of the UCSD basemap. The panel design and functionality are identical to the one shown in Fig. 1. Explore the interactive version at <http://cns.iu.edu/econ-p/hexmap>

highlights all the other subdisciplines linked to it. Use $+$ $-$ in top-left or the scroll wheel on the mouse to zoom, centered on the mouse pointer.

Data overlays use bar graphs per subdiscipline to support the visual exploration and comparison of HITS metrics, see Figs. 2 and 4. Up to four HITS metrics can be selected and compared via these bar graphs. A time slider supports focusing on a specific window of time, such as a period of transformation in a field of interest.

While we tried to make both maps informationally equivalent, there are differences in the features of both visualizations that likely result in slightly different experiences for users. Interactivity tools such as zoom, filters, and details on demand are identical for both. The same goes for the hover functionality that reveals connected nodes. However, when a

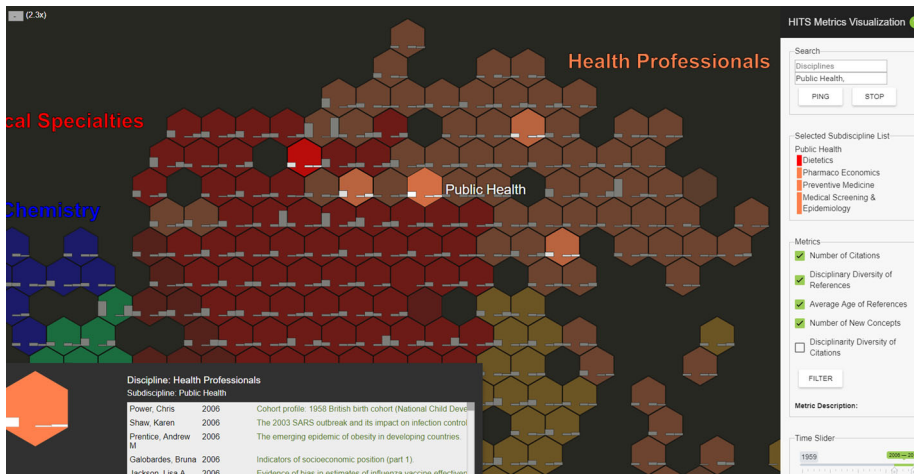


Fig. 4 Hexmap zoomed into the *Health Professionals* discipline. Hovering over the *Public Health* subdiscipline highlights all of the subdisciplines linked to it and lists their names in the right panel. Clicking on it brings up a window at the lower left that lists publications with the highest combined metrics score of the four selected HITS metrics

user hovers over a node in the network-based map, both the text label of the hovered-over node as well as the labels of the connected nodes are displayed while in the hexagon-based visualization, only the text label of the hovered-over hexagon is displayed, see Fig. 5.

This difference is due to the fact that in the hexagon map, unconnected hexagons are dimmed, while in the network map, unconnected nodes are also dimmed, which, however, does not allow the connected nodes to truly stand out. The hexmap leverages the interactive functionality of the web. The UCSD map excels in print usage because all data is present at a single glance.

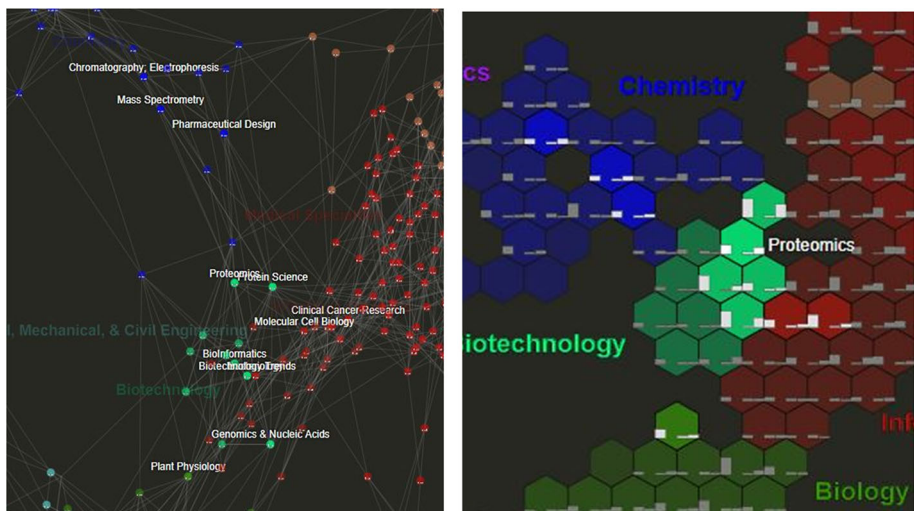


Fig. 5 “Proteomics” node in a hovered state in the network-based map. Note the additional text labels of the connected nodes (left). When tiles in the hexmap are hovered over, linked subdisciplines are highlighted (right)

The visualization of individual metrics differs slightly as well. When a user chooses only one metric to be displayed, in the hexagon-based map, one bar graph appears in each hexagon to indicate the value of the metric. If two or more are selected, additional bars are added. In the network-based map, with one metric selected, its value is encoded by the area size of the node. With two or more metrics selected, the same bar graph visualization is added like in the hexagon-based map, and all nodes are of equal size. In the experimental setup, two or more metrics are selected to make the two visualizations more comparable.

Clicking on a node brings up a lower-left window with the discipline and subdiscipline name, a listing of key papers published in that subdiscipline, and a bar graph visualization of the metrics, reading from left to right in the top-down order of the selected metrics in the right-hand bar. Clicking on a paper title allows the user to read the paper from MEDLINE.

To facilitate searching specific subdisciplines in this rather complex visualization, we implemented a Ping functionality. In the upper-right corner of the visualization, users can type in disciplines and subdisciplines for which they are looking. For example, if one were to type in “drug”, the users would be presented with 2 suggestions: “Drug Safety” and “Drug Discovery” (see Fig. 6, left). Selecting either, and then clicking the Ping button (see Fig. 6, right), will cause the node or hexagon to start blinking (see Fig. 7). This functionality is not compared in the user study.

Administrative interface for metrics management

An administrative interface supports the upload of revised or new metrics and the re-computing of science map visualizations. It is intended for use by researchers and metrics developers and a snapshot of the interface is given in Fig. 8. Existing metrics can be enabled or disabled in the UCSD network and hexmap visualization interfaces. New metrics can be added by entering a metric identifier (ID), a metric name, and brief description and uploading a file that lists the new metric’s value for each publication (PubMed ID). All metrics for which Display is “yes” show in both visualization interfaces.

Science map comparison

This section details the comparison of the UCSD network and hexmap interactive interfaces introduced in the previous section. We present our research hypotheses first, then describe the experimental and data logging setup, and last but not least detail the user data analysis and results.

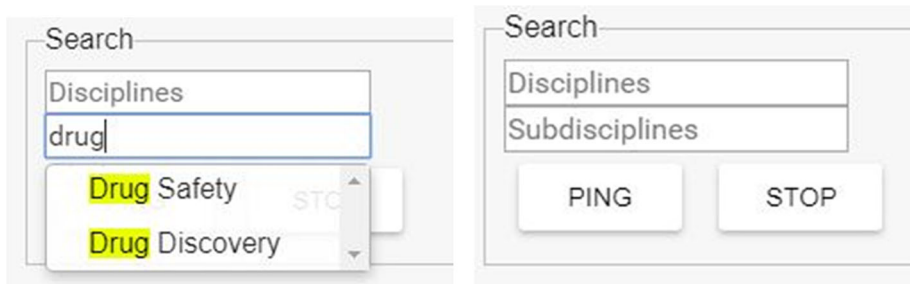


Fig. 6 The search functionality with suggestions (left) and Ping and Stop buttons (right)

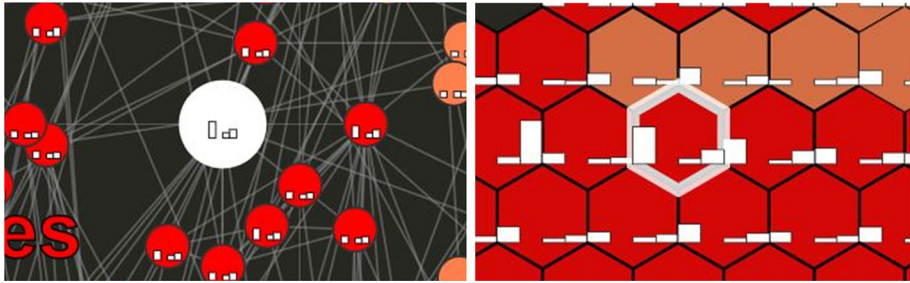


Fig. 7 *Drug Safety* node when pinged in network (left) and hexmap (right)

HITS Metrics Visualizations: Administrative Interface

This interface supports the management of existing and upload of new high-impact and transformative science (HITS) metrics [1] for the HITS Metrics Visualizations, see [ECON-UCSDMap](#) and [ECON-HexUCSDMap](#).

Modify Existing Metrics

To make an existing metric (in)visible in the HITS Metrics Visualization, simply click its pen-shaped Edit icon in the table below, and toggle True/False for Display. To delete an existing metric, click on (-) in Remove column.

Metric ID	Metric Name	Metric Description	Display	Edit	Remove	Upload Data
2	Number of New Concepts	The number of unique top concepts that an article uses within 5 years of the concept's vintage.	Yes	...	...	Choose File No file chosen Upload
3	Average Age of References	Average difference in age between an article and the articles that it cites.	Yes	...	...	Choose File No file chosen Upload
4	Number of Citations	Number of citations of an article within articles indexed in MEDLINE up until 2014.	Yes	...	...	Choose File No file chosen Upload
5	Disciplinary Diversity of Citations	Herfindahl index of an article based on the field of forward citation articles.	Yes	...	...	Choose File No file chosen Upload

Add New Metric

To add a new HITS metric, enter Metric ID, Metric Name, and Metric Description into the blank fields:

Metric ID	Metric Name	Metric Description	Display ☐	CREATE
-----------	-------------	--------------------	----------------------------------	---------------------------------------

Check Display if metric should be visible in the HITS Metrics Visualization. Select CREATE to add new metric. The new metric will then appear in the table of existing metrics (see above). Next, upload metrics data via Choose File. The data needs to be in a comma-separated format and list the HITS metric value for each PubMed ID, see [sample file](#).

After all changes have been made, click COMPUTE (below) to update metrics and compute new ranking values.

Fig. 8 Administrative interface for metrics developers in support of metrics management

Research hypotheses

Ultimately, we are interested in comparing the two science map interfaces—network versus hexmap—in terms of four basic features:

- **Legibility** Is the map readable (e.g., can users find certain disciplines, subdisciplines, links, paths, clusters?) using timed tasks (e.g., finding a certain discipline, identifying the subdiscipline with the highest impact), exploratory tasks (e.g., understanding spatial relationships between disciplines), and targeted tasks (e.g., identifying specific subdisciplines that interlink disciplines).
- **Memorability** Can participants remember node and link attributes (e.g., the placement of disciplines and their interconnectivity) and network attributes (e.g., names and number of paths and clusters, overall structure and shape) in the short term (e.g., minutes) and in the long term (e.g., days and weeks).

- *Engagement* How much time do users spend with each visualization? How much do they like each visualization?
- *Utility* Which science map provides the most actionable insights and/or prompts more meaningful questions, or is otherwise useful for human understanding and decision-making?

In this paper, we focus on four hypotheses:

H1: Spatial Ability Impact Performance Users that score higher on a spatial ability test will be able to (a) better memorize the position of the text labels in a screenshot of the visualization and (b) to compare metrics more easily.

H2: Classification Due to the great prevalence of network visualization, we predict that more users will be able to correctly classify their assigned visualization in the network than hexmap condition.

H3: Memorization of Discipline Positions We postulate that the hexmap makes it easier for users to memorize the position of scientific disciplines when compared to the network map. That is, after examining a static version of the map, users will be able to place discipline labels with higher accuracy.

H4: Utility for Metrics Comparison The hexmap will make it easier than the network visualization to retrieve metrics and compare them between subdisciplines.

The experimental setup aims to test each of the four hypotheses and is discussed next.

Experimental setup

Qualtrics (<http://cns.iu.edu/docs/data/2018-ECON-SciMaps/econ-qualtrics-survey.pdf>) was used in combination with a jspsych site (<http://www.jspsych.org/plugins/jspsych-free-sort/>) (de Leeuw 2015) to collect data on the spatial layout of discipline labels. The setup of this study conforms to standard human subject study procedures. After obtaining informed consent, participants completed five steps:

1. *Pre-test questionnaire* that captured information on job title, department, student status, age, gender, and native language. Subsequently, subjects were asked about their familiarity with data visualizations (e.g., “How comfortable are you reading tables, charts, and graphs?”), and whether or not they ever took an information visualization class.
2. *Spatial ability test* where each participant had to solve a set of five problems, including folded cube and stacked cubes, see examples in Fig. 9. The tests were taken from <http://www.psychometric-success.com/aptitude-tests/spatial-ability-tests-cubes.htm> and <https://www.wikijob.co.uk/content/aptitude-tests/test-types/spatial-awareness-test>. The data collected here was used to test **H1(a)** and **H1(b)**.
3. *Memorization task* Next, subjects were randomly assigned to two cohorts, ‘hexmap’ and ‘network’, depending on which visualization they were presented with. Each cohort was asked to explore a static picture of their respective visualization for up to 3 min. A static screenshot was used to ensure that each subject saw the very same visualization without zoom/pan, filter, or details on demand interactivity. To increase legibility in the Qualtrics setup, two modifications were made: white outlines were added to text labels, and the positions of the text labels were slightly altered for better readability (see page 11 in Qualtrics survey ([!\[\]\(6c2f23024eecffbc3f38b5f281930737_img.jpg\) Springer](http://cns.iu.edu/docs/data/2018-ECON-

</div>
<div data-bbox=)

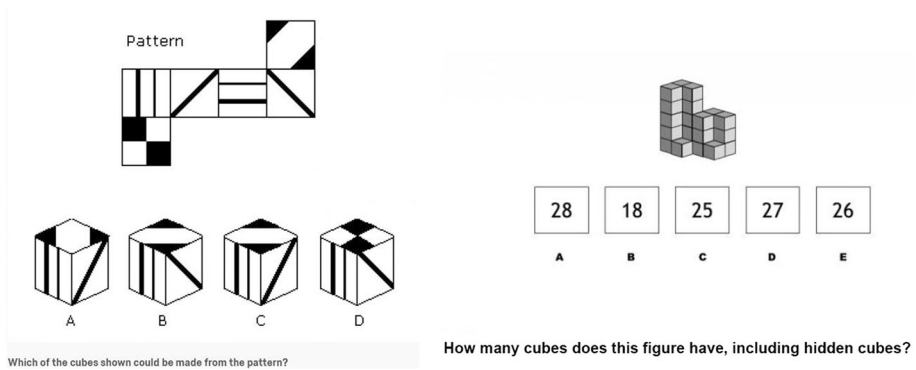


Fig. 9 Folded cube spatial ability task (left) and stacked cube task (right)

[SciMaps/econ-qualtrics-survey.pdf](#)). After subjects spent up to 3 minutes memorizing the positions of the text labels, the screen changed into an interactive jspsych-supported site with the 13 discipline labels in random placement, see Fig. 10. Subjects were asked to move the discipline labels to their positions in the original map. The final placement of fields was captured and automatically sent to a MySQL database for data logging in order to test **H3**.

4. *Classification and retrieval tasks* Subsequently, subjects were asked to explore the interactive visualizations introduced in “[Science basemap construction, data overlays, interactivity](#)” section. The ‘hexmap’ cohort got the hexmap visualization, the ‘network’ cohort got the network visualization. The participants were asked to classify the visualization as network or a map, identify subdisciplines with specific metric values, and access details on demand. We used this part of the experiment to test **H2** and **H4**.

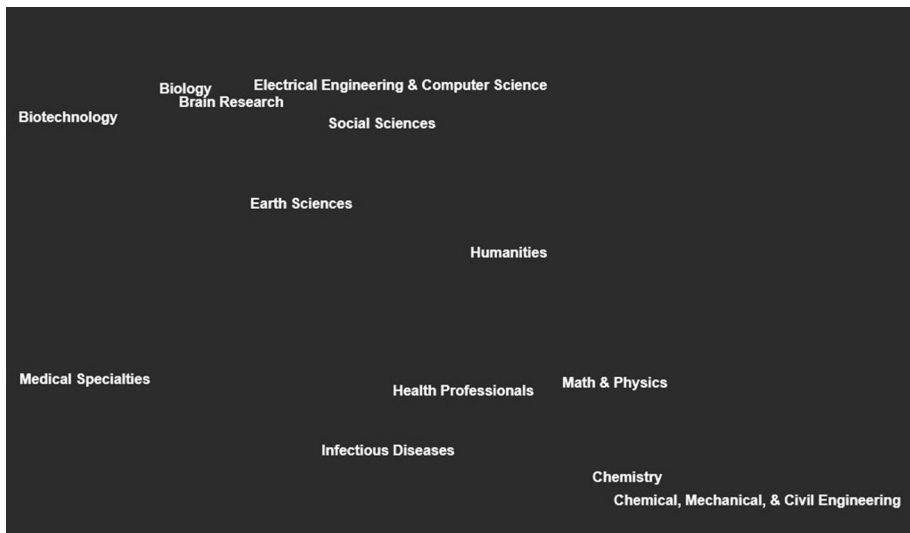


Fig. 10 Interface for the spatial arrangement of disciplines task. Notice the randomized position of non-color coded discipline labels

5. *Post-questionnaire* After the main experiment, an on-line questionnaire was given that contained open-ended questions such as “What did you like about the visualization?” to gather additional quantitative data using Likert scales and qualitative feedback. The primary goal of these questions was to improve the visualization for future users.

All study instruments are available online:

Visualization 1 (hexmap) (<http://cns.iu.edu/econ-p/hexmap>)

Visualization 2 (network) (<http://cns.iu.edu/econ-p/ucsdmap>)

jpspsych spatial arrangement test (<https://www.indiana.edu/~pcl/robsexperiments/tests&examples/fieldArrange.html>)

PDF of full Qualtrics survey: (<http://cns.iu.edu/docs/data/2018-ECON-SciMaps/econ-qualtrics-survey.pdf>)

p. 1–3: Study information sheet

p. 3–5: Pre-questionnaire

p. 6–10: Spatial ability test

p. 10–12: Instructions for spatial placement part

p. 12–17: Questionnaire with tasks

p. 18–21: Post-questionnaire

In order to match data from the spatial placement task with data from Qualtrics, each participant’s unique ResponseID was passed from Qualtrics to the jpspsych site through a query string in an auto-generated URL given to each participant to uniquely identify each user and facilitate matching of both datasets. The position of the final (x, y) position for each discipline string was logged using a JSON object array (recording the (x, y) positions of all 13 disciplines for each participant).

Participants from the Psychological and Brain Sciences’ (PBS) Course Credit Subject Pool at Indiana University (<https://iub-pbs-credit.sona-systems.com>) were invited to participate in the experiment. Note that this restricts the survey to students enrolled in approved, introductory-level PBS courses. Students did receive Participation Course Credit for starting the experiment.

Data processing and analysis

Exactly 76 participants (all but 4 were students) took the study. Data from 29 participants had to be discarded as they did not complete all parts of the experiment, leaving a sample of **47 subjects—26 for the hexmap layout and 21 for network layout**. Qualtrics and jpspsych data for all participants were retrieved as *.csv tables, and ResponseIDs were matched to create 47 full datasets.

From the pre-questionnaire, we gathered that all 47 participants were students (30 freshmen, 8 sophomores, 5 juniors, 3 seniors, 1 unknown). Ages ranged from 18–20 (41), 21–30 (5), to 31–40 (1). Study subjects included business (8), social work (1), nursing (1), finance (3), fundraising (1), university diversity (1), human biology (4), contemporary dance (1), computer science (2), golf (1), media or media advertising (5), arts and sciences (1), sociology (1), economics (1), psychology (1), and others or unknown (15). 28 participants identified as female, 18 as male, and 1 participant preferred not to reveal their gender. 36 participants were primarily English native speakers, 9 grew up speaking Asian languages, and 2 have Portuguese as their first language. Only 3 of the participants had ever taken data visualization training. When asked how comfortable they were reading tables, charts, and graphs, 5 said they were “very uncomfortable”, 10 were “somewhat

uncomfortable”, 7 were neither of the two, 19 were somewhat comfortable, and 6 were very comfortable.

The spatial ability test asked each participant to complete five tasks. Only 1 participant was able to answer all questions correctly, 9 provided 4 correct answers, 18 provided 3 correct answers, 12 provided 2 correct answers, 6 gave 1 correct answer, and 1 subject got all questions wrong.

Using Processing (<https://processing.org/>) and R (<https://www.r-project.org/>), we extracted matrices of the relative positions of the labels as placed by participants from their memory, as well as the matrices for the positions of the labels in the screenshot of the visualization they used for the memorization task. The relative distance matrix for the hexmap visualization is shown in Table 1. With Processing and R, we computed a correlation coefficient for each participant to tell us how similar the relative distances of the labels as placed by the participant are to the original distances. Comparing these matrices yielded a list of correlation coefficients we could use for further data analysis.

Using R, we found a correlation of **0.314** between spatial ability score and relative distance correlation coefficient for the hexmap cohort, and **-0.002** for the network group. While these are not statistically significant per se, these numbers could indicate a weak positive correlation between the spatial ability score hexmap participants get and the fidelity of their relative placements in the spatial arrangement test. That is, the higher subjects scored, the more accurate they placed their labels. It is quite evident that there is no correlation between those two metrics for the network cohort, which indicates that spatial reasoning ability might have little influence on the ability to memorize and place text labels correctly. Given that we do not have a statistically significant correlation for the hexmap cohort, we **fail to confirm H1(a)**, which stated that users who score higher on a spatial ability test will be able to better memorize the position of the text labels in a screenshot of the visualization. In **H1(b)**, we postulated that subjects with a higher score in the spatial ability test would score higher on the questionnaire tasks where they had to identify disciplines with the highest metrics and explore the visualization to gather information about the total number of disciplines and subdisciplines. We arrived at a correlation of **0.087** for hexmap, and **-0.16** for network, which does not allow us to reject the null hypothesis of no influence of spatial ability test scores and performance on metrics comparison and identification. One reason for this might be the limited content of the

Table 1 Matrix of relative distances between original labels in hexmap visualization

	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	
L0		0	3412.517	564.257	670.4364	2232.052	2855.874	2292.72	1729.778	1226.511	2332.1	1376.611	2292.643	3684.899
L1	3412.517		0	3239.128	3295.526	1245.492	1170.961	1660.302	1731.074	2806.347	1357.933	2076.083	2076.256	999.3083
L2	564.257	3239.128		0	1167.533	2163.775	2518.029	1912.788	1695.728	704.3919	2022.988	1167.117	1832.869	3366.571
L3	670.4364	3295.526	1167.533		0	2054.001	2971.359	2494.442	1566.082	1704.148	2439.098	1478.935	2606.929	3744.634
L4	2232.052	1245.492	2163.775	2054.001		0	1343.046	1307.875	504.3659	1919.066	993.443	1032.981	1720.625	1874.51
L5	2855.874	1170.961	2518.029	2971.359	1343.046		0	629.877	1572.384	1927.765	533.2954	1506.411	958.358	851.2227
L6	2292.72	1660.302	1912.788	2494.442	1307.875	629.877		0	1320.556	1300.105	346.902	1020.699	443.8119	1476.932
L7	1729.778	1731.074	1695.728	1566.082	504.3659	1572.384	1320.556		0	1552.244	1096.148	649.6191	1655.312	2243.495
L8	1226.511	2806.347	704.3919	1704.148	1919.066	1927.765	1300.105	1552.244		0	1486.755	905.3469	1147.325	2776.96
L9	2332.1	1357.933	2022.988	2439.098	993.443	533.2954	346.902	1096.148	1486.755		0	973.7603	790.3834	1354.155
L10	1376.611	2076.083	1167.117	1478.935	1032.981	1506.411	1020.699	649.6191	905.3469	973.7603		0	1209.036	2316.055
L11	2292.643	2076.256	1832.869	2606.929	1720.625	958.358	443.8119	1655.312	1147.325	790.3834	1209.036		0	1744.663
L12	3684.899	999.3083	3366.571	3744.634	1874.51	851.2227	1476.932	2243.495	2776.96	1354.155	2316.055	1744.663		0

spatial ability test, and the unsatisfactory statistical strength that comes with it. We **fail to confirm H1(b)**.

Memorization task As explained above, we used Processing (<https://processing.org>) to visualize data from the spatial arrangement test and to calculate the relative distances between each participant's discipline labels. This yielded a matrix as seen in Table 1, and comparing each participant's matrix to the one with the relative distances of the original visualization screenshot the subjects used to memorize the positions, we calculated a correlation coefficient for each subject, with 0 being no similarity whatsoever, and 1 being a complete match. In order to visualize the placements of the subjects in relation to the original screenshots, we took the position of the text labels in the screenshots presented to the participants and calculated a stretch factor variable for each participant so all data could be displayed in one 1200 x 600 pixels window, the same resolution used for the spatial arrangement test on the aforementioned jspsych website (see Fig. 10). For example, the screenshot of the network map had a resolution of 4468 x 2190 pixels (displayed smaller in Qualtrics but with the same aspect ratio). We measured the label positions on this screenshot, and then divided $w(\text{participant})/w(\text{original})$ and $h(\text{participant})/h(\text{original})$, where w and h are variables for width and height. This gave us stretch factors of 0.269 and 0.274 for width and height, respectively. We multiplied the measured label positions from the screen shots by these stretch factors and included them in our Processing visualizations (see the squares in Figs. 11, 12) so that the participants' label positions were comparable to the original label positions they had to memorize. White lines were added to show the distance between participant-placed label (circles) and correct position (squares). As can be seen from the figures, the positions of labels on the outer side of the visualization were easier to memorize than the ones closer to the middle. For example, the *Math & Physics*, *Electrical Engineering & Computer Science*, and *Chemical, Mechanical, & Civil Engineering* labels, all positioned toward the outer edge of the actual map, are all closer to each other than the other ones, in both visualizations. A possible explanation could be that participants spent more time memorizing the outermost labels, or that memorizing these is

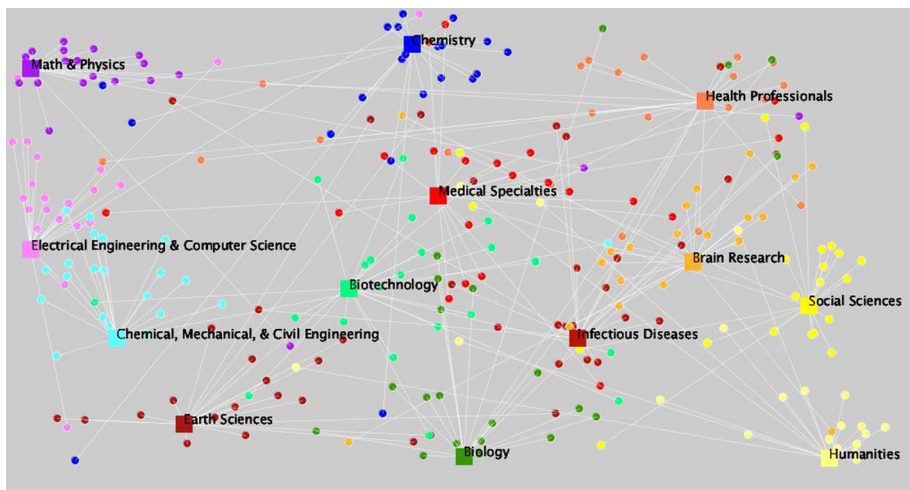


Fig. 11 Map of original label positions (squares) and labels placed by human subjects (circles) for the network map, with white lines visualizing the distance and direction of each label towards its correct position

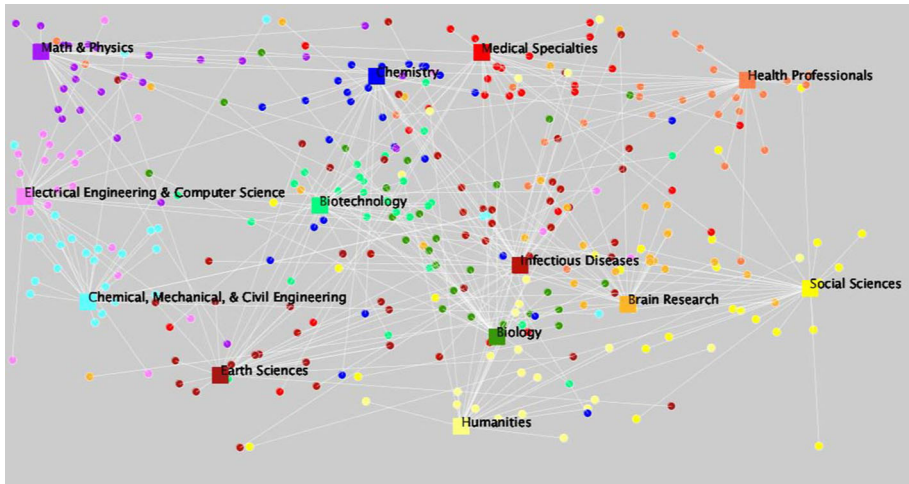


Fig. 12 Map of original discipline label positions (squares) and labels placed by subjects (circles) for the hexmap, with white lines visualizing the distance and direction of each label towards its correct position

simply easier. Alternatively, there is a well-known influence by which spatial judgments are anchored to salient categorical information, and the display edges very likely serve as such anchor points for participants' spatial judgments, with participants using internal codes such as “*Electrical Engineering* was close to the upper-left hand corner of the display”. Also, on both the hexmap and the network visualization, the larger “landmasses” are on the right side, leading to clutter, potentially making it hard to memorize the exact location of labels. For a boxplot of the relative distance matrix coefficients of both cohorts, see Fig. 13, left.

Each participant was suggested to take 3 minutes to memorize the location of the labels in the screenshot of the visualization. We recorded how much time they actually spent on this, and then computed the correlation between time spent memorizing and the relative distance matrix correlation coefficient of each participant; the results are **-0.002** for the network, and **0.4** for the hexmap visualization. This finding is illuminating, because while

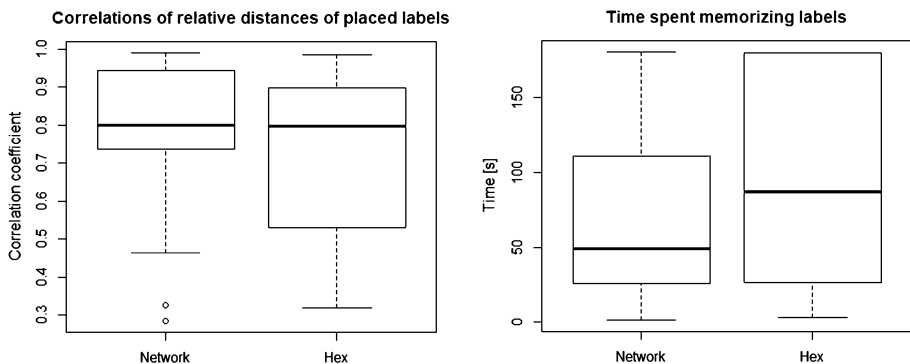


Fig. 13 Boxplot of relative distance matrix correlation coefficient between reported labels for each cohort (left) and time spent memorizing original text label positions for each cohort (right)

there is a moderate positive correlation between time spent memorizing and the label placement for the hexmap, no such correlation seems to exist for the network visualization. In this experiment, a positive correlation might be expected under the assumption that spending longer to study the visualization would result in more accurate positioning (e.g. a more accurate matrix). What this could indicate is that the hexmap might actually make it easier for the participants to memorize the labels, while for the network map, it makes no difference how long the participants viewed it. For a depiction of how much time both cohorts spent memorizing, see Fig. 13, right.

Additionally, we performed a **t-test** to see if we can find statistically significant differences between the two groups with regards to relative distances. We obtained a p-value of **0.49**, and consequently do **not confirm H3**, which stated that the hexmap makes it easier for users to memorize the positions of scientific disciplines when compared to the network map. Fig. 13, left shows boxplots of the relative distance matrix correlation coefficients.

Classification and retrieval tasks After having completed the spatial arrangement task outside of Qualtrics, the subjects returned to the survey and had to perform a set of tasks, retrieving information from the visualization. Having opened the actual, interactive visualization, they were asked to determine the type of the visualization they were shown. 14 out of 21 members of the network cohort answered correctly ‘network’, and 16 out of 26 hexmap subjects correctly answered ‘map’. We performed a X^2 test between the two groups, and got a p-value of **0.953** forcing us to **fail to confirm H2**, which stated that a higher number of network visualization users will be able to classify their visualization as a network than the number of hexmap users that will be able to classify their visualization as a map.

Asked which subdiscipline had the highest number of new concepts (“Drug Discovery”), only 4 participants (2 from each cohort) figured out the solution. However, 7 members of network and 1 of hexmap answered not with the correct subdiscipline but the correct discipline, meaning they probably saw the size of the biggest node (network) or the little bar graph (hexmap) visualizing the metric. It was just a problem to click on or hover over the actual node to retrieve the subdiscipline name, or the question was only partly understood by the subject. Asked which subdiscipline out of 3 had the highest number of new concepts (“Endoscopic Surgery”), 24 participants (10 network, 14 hexmap) answered correctly.

When a node or hexagon is clicked, a legend appears at the bottom, containing bibliographic information about papers ranking highly on the metrics of the particular subdiscipline. When asked to explain what information the legend at the bottom of the screen showed, only 20 participants provided a substantial answer. While participants did not know how to parse this bibliometric data, most of them seemed to be able to at least retrieve it. When asked for the first name of the first author on the paper with the highest “Number of Citations” metric in the “Thermal Analysis” subdiscipline, 18 participants (8 network, 10 hexmap) were able to provide a satisfactory answer.

Finally, participants were asked to estimate the number of disciplines and subdisciplines; they were encouraged to explore the visualization and its context rather than counting; the information was retrievable by clicking on the “?” sign in the top right corner that brings up information about the visualization, stating that there are 554 subdisciplines and 13 disciplines. Only 9 participants correctly determined the count for disciplines and subdisciplines; 6 gave only the number of subdisciplines.

To find out more about how the two visualizations empowered subjects to retrieve metrics-related information, we performed a X^2 test for differences between the two cohorts, and got a p-value of **0.106**, meaning that we **fail to confirm H4**, which stated that

the hexmap will make it easier than the network visualization to retrieve metrics and compare them between subdisciplines. However, it seems desirable to continue testing for statistical significance of differences in performance of both cohorts with larger sample sizes.

Post-questionnaire data included open-ended and Likert-scale answers. Asked how much they liked the visualization, a total of 22 subjects answered “not at all” (7 from network, 15 from hexmap), 15 participants said “somewhat” (10 network, 5 hexmap), 6 “neither liked it nor disliked it” (2 network, 4 hexmap), and 4 subjects did not answer this question. Furthermore, out of our 47 subjects, 17 said the visualization was “not easy at all” (5 network, 12 hexmap), 17 said “not very easy” (10 network, 7 hexmap), 1 network subject said “neither easy nor hard”, 8 said it was “somewhat easy (3 network, 5 hexmap), and 4 did not answer the question.

In terms of what participants specifically liked about their respective visualizations, colors were mentioned by members of both cohorts, as was their value for encoding complex data (“easier for broader look”, “neat way to organize information”). What becomes clear from looking through this qualitative feedback is that most participants felt comfortable with the visualization as long as they did not have to retrieve information requiring them to dig deeper. The macro-view of the visualization allowed many of them to gain macro-scale insights (“It’s colorful and looks cool. It helps the audience see connections and relationships between fields”). On the down side, members of both cohorts found their visualization “confusing”, and hard to “navigate” and “search for stuff”. What this means is that most subjects were probably comfortable with using the visualizations in a static (or minimally interactive) manner, like zooming and panning a little bit, without manipulating the displayed data or retrieving node-specific information.

Discussion and outlook

Current results do not show significant differences in how the two maps were remembered or used by participants, except that hexmap users spent a lot more time looking at and memorizing the labels on the hexmap.

As a general takeaway from this experiment, it seems that the layout of the basemaps was generally more easily understood than the built-in functionalities such as subdiscipline search and metrics filters. The derivative functions add considerable complexity to the visualizations, adding to the general challenge of the visualizations which are largely unfamiliar to most audiences, especially undergraduate students who had no prior visualization training, and who can be assumed to have very limited exposure to science maps or impact metrics. Given the difficulties that our novice subjects had with using the visualizations to derive insights, we are planning to conduct a user study with expert users on domain specific maps, e.g., featuring Alzheimer research. For the future development of interactive online visualizations aimed at amateur audiences, we recommend to a) design the visualization in a way that allows for insights even if the user does not interact much with the visualization b) keep up the good practice of using color hue for encoding qualitative data c) limit the amount of information revealed in each step a user executes as too much new information can lead to confusion and frustration. We hope that the detailed discussion of the experiment setup will inspire other scholars to conduct well-controlled, larger-scale human subject experiments comparing not only static but also online interactive maps of science.

Acknowledgements The authors would like to thank Joseph Staudt, Huifeng Yu, Robert Light, Gerald Marschke, and Bruce Weinberg from the HITS metrics team for their expert comments on the design and functionality of the two visualizations. Michael Frisby and David Endicott from the Social Science Research Commons at Indiana University Bloomington provided data analysis support. Börner, Simpson, and Bueckle are partially supported by the National Institutes of Health under awards P01 AG039347 and U01CA198934 and the National Science Foundation under award EAGER 1566393, NCN CP Supplement 1553044, and AISL 1713567. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This study was accepted by the Human Subjects Office of Indiana University Bloomington as meeting the criteria of exempt research as described in the Federal regulations at 45 CFR 46.101(b), paragraph 2. The Human Subjects Office accepted the use of an information sheet.

References

- Börner, K. (2010). *Atlas of science: Visualizing what we know*. Cambridge, MA: The MIT Press.
- Börner, K. (2015). *Atlas of knowledge: Anyone can map*. Cambridge, MA: The MIT Press.
- Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179–255.
- Börner, K., Klavans, R., Patek, M., Zoss, A., Biberstine, J. R., Light, R., et al. (2012). Design and update of a classification system: The UCSD map of science. *PLoS ONE*, 7(7), e39464. <https://doi.org/10.1371/journal.pone.0039464>.
- Boyack, K. W., Börner, K., & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, 79(1), 45–60.
- Boyack, K. W., & Klavans, R. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, 60(3), 455–476.
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374.
- Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., et al. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS ONE*, 6(3), e18029. <https://doi.org/10.1371/journal.pone.0018029>.
- Chen, C. (2003). *Mapping scientific frontiers*. London: Springer Verlag.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
- Emmons, S., Kobourov, S., Gallant, M., & Börner, K. (2016). Analysis of network clustering algorithms and cluster quality metrics at scale. *PLoS ONE*, 11(7), e0159161. <https://doi.org/10.1371/journal.pone.0159161>.
- Exhibit Curators. (2005). *Places & spaces: Mapping science* exhibit. <http://scimaps.org>.
- Farrugia, M., & Quigley, A. (2011). Effective temporal graph layout: A comparative study of animation versus static display methods. *Information Visualization*, 10(1), 47–64.
- Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 122, 108–111.
- Garfield, E. (1974). ISI's atlas of science may help students in choice of career in science. *Essays of an Information Scientist* 1974–1976, 2(38), 311–314. <http://www.garfield.library.upenn.edu/essays/v2p311y1974-76.pdf>.
- Garfield, E. (1981). *ISI atlas of science: Biochemistry and molecular biology 1979–1980*. Philadelphia, PA: ISI.
- Garfield, E. (1987). Mapping the world of nutrition: Citation analysis helps digest the menu of current research. *Essays of an Information Scientist*, 10, 349. <http://www.garfield.library.upenn.edu/essays/v10p349y1987.pdf>.
- Garfield, E. (1988). Mapping the world of epidemiology: Part 2. The techniques of tracking down disease. *Essays of an Information Scientist*, 11, 290. <http://www.garfield.library.upenn.edu/essays/v11p290y1988.pdf>.
- Garfield, E., & Shea, J. E. (Eds.). (1984). *ISI atlas of science: Biotechnology and molecular genetics 1981/82*. Philadelphia: ISI.
- Ghani, S., & Elmqvist, N. (2011). Improving revisitation in graphs through static spatial features. In *Proceedings of Graphics Interface*, pp. 737–743.
- Saket, B., Scheidegger, C., Kobourov, S., & Börner, K. (2015). Map-based visualizations increase recall accuracy of data. In *EUROGRAPHICS 2015, Zürich, Switzerland*, pp. 441–450.

- Saket, B., Simonetto, P., Kobourov, S., & Börner, K. (2014). Node, node-link, and node-link-group diagrams: An evaluation. *IEEE Visual Analytics Science and Technology, IEEE Information Visualization, and IEEE Scientific Visualization, Paris, France*, 20(12), 2231–2240.
- Shiffrin, R. M., & Börner, K. (2004). Introduction: Mapping knowledge domains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1), 5183–5185.
- Skupin, A., Biberstine, J. R., & Börner, K. (2013). Visualizing the Topical structure of the medical sciences: A self-organizing map approach. *PLoS ONE*, 8(3), e58779. <https://doi.org/10.1371/journal.pone.0058779>.
- Staudt, J., Huifeng, Yu., Light, R., Marschke, G., Börner, K., & Weinberg, B. (2016). *High-Impact and Transformative Science (HITS) metrics*. Ghent, Belgium: OECD Blue Sky III.