Toward the Operationalization of Visual Metaphor

Alexis Hiniker*
Human Centered Design and Engineering, Box 352315, University of Washington, Seattle, WA 98195. E-mail: alexisr@uw.edu

Sungsoo (Ray) Hong*
Human Centered Design and Engineering, Box 352315, University of Washington, Seattle, WA 98195. E-mail: rayhong@uw.edu

Yea-Seul Kim
Human Centered Design and Engineering, Information School, Box 352840, University of Washington, Seattle, WA 98195. E-mail: yeaseul1@uw.edu

Nan-Chen Chen
Human Centered Design and Engineering, Box 352315, University of Washington, Seattle, WA 98195. E-mail: nanchen@uw.edu

Jevin D. West
Information School, Box 352840, University of Washington, Seattle, WA 98195. E-mail: jevinw@uw.edu

Cecilia Aragon
Human Centered Design and Engineering, Box 352315, University of Washington, Seattle, WA 98195. E-mail: aragon@uw.edu

Many successful digital interfaces employ visual metaphors to convey features or data properties to users, but the characteristics that make a visual metaphor effective are not well understood. We used a theoretical conception of metaphor from cognitive linguistics to design an interactive system for viewing the citation network of the corpora of literature in the JSTOR database, a highly connected compound graph of 2 million papers linked by 8 million citations. We created 4 variants of this system, manipulating 2 distinct properties of metaphor. We conducted a between-subjects experimental study with 80 participants to compare understanding and engagement when working with each version. We found that building on known *image schemas* improved response time on look-up tasks, while contextual detail predicted increases in persistence and the number of inferences drawn from the data. Schema-congruency combined with contextual detail produced the highest gains in comprehension. These findings provide concrete mechanisms by which designers presenting large data sets through metaphorical interfaces may improve their effectiveness and appeal with users.

*Alexis Hiniker and Ray Hong contributed equally to this paper.

E-mail: alexisr@uw.edu, rayhong@uw.edu, yeaseul1@uw.edu, nanchen@uw.edu, jevinw@uw.edu, aragon@uw.edu

Received October 8, 2016; revised December 9, 2016; accepted December 21, 2016

 • Published online 7 August 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23857

Introduction

In its modern conception, metaphor is understood to be much more than a rhetorical device or feature of poetic language (Hamilton, 2000b). Contemporary theories of metaphor explain that it is, rather, a way of thinking, and one so natural that it is largely automatic and unconscious (Gentner, 1983; Lakoff, 1993). Metaphor enables us to align novel concepts with ones we already understand and is a common mechanism for assimilating new knowledge (Blackwell, 2006; Lakoff, 1993). Prior work shows that thinking metaphorically is instrumental to successful scientific and creative processes (Blanchette & Dunbar, 2001; Dunbar, 1995), and that the ability to think in metaphor has been essential to numerous significant inventions (Gordon, 1974). More modestly, but at least as important, metaphor is embedded in much of our everyday thought and enables routine cognitive off-loading (Fischer, Itoh, & Inagaki, 2014; Lakoff, 1993).

Designers have long been interested in using metaphorical representations to provide users with cognitive scaffolding. By casting an interface in terms of a familiar concept, a designer can facilitate easy access to interface features or the data behind them (Blackwell, 2006). For example, by telling a user that a word processor contains "cut," "copy," and "paste" functionality, a designer can catalyze a user's understanding of these features in a way that would be more laborious without the assistance of a well-understood concept.

This technique is particularly relevant to the visualization of large and complex data sets, where the vast number of data points, complexity of relationships, and high dimensionality all introduce an abundance of detail that is unmanageable without cognitive support. As a user cannot possibly process all aspects of the data in their raw form, one of the fundamental challenges of information visualization is transforming these large quantities of information into visual structures that the viewer can interpret meaningfully (Thomas & Cook, 2006). Metaphor, a workhorse of cutting through surface detail to reveal underlying structure and structural relationships, offers to serve as a promising tool for designers in this space. Yet an evidence-based understanding of how designers might strategically incorporate metaphor into user interfaces remains elusive. Some interfaces built on visual metaphor have been wildly successful, while others have fallen flat, and the design community has yet to provide a clear rationale for this disparity (Hamilton, 2000b).

In this paper we present an original data visualization and manipulate it to evaluate two different elements of metaphor. We operationalize interface metaphor in terms of: i) its congruency with documented *image schemas* (structured patterns of thinking which enable metaphorical thought), and ii) visual embellishment reflecting the surface detail of the metaphor. We hypothesized that image-schema alignment would improve the success of visual metaphor based on prior work showing its effectiveness in guiding the design of traditional user interfaces. We hypothesized that visual embellishment would also improve the effectiveness of visual metaphor by making the metaphor more apparent and facilitating memorability.

To evaluate these hypotheses, we extracted a large corpus of data and built an interactive visualization enabling users to explore it. Our visualization tool presents the set of academic fields that emerge organically from the JSTOR corpora of 2 million scholarly articles connected hierarchically by 8 million citations. The scientific literature is a large, growing, and complex graph connected through its citations and footnotes. The need for improved interfaces to explore this network and the relative sparsity of effective visualization methods and theory for navigating these complex systems motivated our investigation.

We present results from an evaluation with 80 participants, comparing four different versions of our tool. Our results suggest that choosing a metaphor that is an appropriate structural representation of the data set it presents and surfacing this metaphor with visual detail can improve users' understanding of and willingness to engage with the data.

Background and Related Work

Interface Design and Metaphor

Metaphor has had a history of mixed success in interface design. For example, the desktop metaphor is heralded as a well-received and highly effective interface (Barr, Khaled, Noble, & Biddle, 2005). Other mainstream successes grounded in metaphor permeate daily life: the familiar cut-copy-paste tools; the word-processor-as-typewriter interface (including "carriage returns" and a ruler); and the short burst of a songbird "tweet" representing a communication limited to 140 characters. However, Apple's use of a trashcan metaphor for disk-ejection is remembered first and foremost for the "horror" and confusion it caused its users (Hamilton, 2000a). Microsoft Bob and Apple's Magic Cap, metaphorical representations of the operating system, were both metaphorical interfaces that were crticized by experts and struggled commercially (Gentner & Nielsen, 1996). In the research community, empirical investigations of metaphorical interfaces have in some instances failed to see predicted performance gains (Ziemkiewicz & Kosara, 2008) and in others even been detrimental to user performance (Blackwell & Green, 1999). Defining the elements that make visual metaphor effective is not a straightforward endeavor (Blackwell, 2006; Blackwell & Green, 1999; Santos et al., 2000).

The use of visual metaphors has also been the subject of research in information visualization. Research teams have presented object-oriented classes as celestial bodies in orbit (Graham, Yang, & Berrigan, 2004), search results as flowers (Chau, 2011), and data feeds as layers of sediment (Huron, Vuillemot, & Fekete, 2013). These investigations suggest that well-chosen metaphors can enhance the effectiveness of visualizations of large data sets but are difficult to replicate or generalize, motivating our investigation to identify specific ways in which metaphors provide support to users. Traditional presentations of relational data sets are also informally grounded in intuitive visual metaphors. Trees representing hierarchies have "roots" and "branches," and through metaphorical pluralism, the process of using multiple metaphors to describe a single concept in order to capture different components of its underlying structure (Hamilton, 2000b), seamlessly shift into having "parents," "children," and "siblings." Here we extend prior work in formalizing the use of metaphor in interface design into the domain of information visualization.

Theoretical Construction of Metaphor

In contemporary linguistics, one prevailing theory of metaphor, advanced by George Lakoff and Mark Johnson, explains that our understanding of the world is built on top of image schemas: rules about the structure of objects, ideas, and relationships that are grounded in bodily experience

(Lakoff, 1993; Lakoff & Johnson, 1997). These schemas allow us to rapidly understand a novel concept by aligning it with a familiar one that shares the same underlying structure.

Lakoff and Johnson argue that these schemas can be extracted from everyday language and the way we speak about concepts. For example, we have a conceptual understanding that ARGUMENT IS WAR.[1] an image schema that is evident in expressions like "engaging in argument," "shooting down her claim," or "attacking his line of reasoning." While these varied metaphorical statements might each appear to stand independently, they all draw from a deeper conceptualization that argument and war have analogous underlying structures and that an understanding of the attributes of one can lead to accurate inferences about the attributes of the other. Our internalized catalog of schemas is extensive (Lakoff, 1994) and, though influenced by our sociocultural environment, is largely grounded in bodily experiences that transcend culture, such as forward motion, occlusion, containment, sensation, and perception (Boers, 1999; Löffler, Lindner, & Hurtienne, 2014).

Although schema-alignment has been discussed in the HCI community (Blackwell, 2006), it is only in recent years that it has gained momentum in its adoption as the theoretical basis for user interfaces. In 2007, Hurtienne and Blessing conducted the first user tests of schema-aligned interface elements and demonstrated these to be more intuitive to users than interface elements that broke with validated schemas (such as a vertical slider with the largest value at the bottom and the smallest value at the top, which breaks with the well-documented schema that MORE IS UP) (Hurtienne & Blessing, 2007). Loeffler and colleagues expanded this work into a design process in which interviews with users were analyzed for metaphorical statements. From these statements, designers isolated underlying schemas that formed the foundation of their interface (Loeffler, Hess, Maier, Hurtienne, & Schmitt, 2013).

Prior work has also theorized that these gains in user understanding will extend to the domain of information visualization. For example, Risch argues that the fact that bar charts typically use a vertical axis which increases upward rather than downward (i.e., bars do not typically start at the top of a chart and grow downward) is not arbitrary, but a result of our existing schema that MORE IS UP (Risch, 2008). We build on this foundation by empirically evaluating this theory with a large and complex data set.

Visual Embellishment and Information Visualization

Extending far beyond the scope of interfaces built on metaphor, there has been considerable debate over the effect of visual embellishment in user interfaces generally (Hullman, Adar, & Shah, 2011). The traditional view in information visualization urges designers to minimize superfluous visual detail, presenting users only with nonredundant data information (Tufle, 1983). This position has been corroborated by empirical studies that have shown, for example, that users performing look-up tasks on charts and graphs achieve higher accuracy and faster response times when extraneous visuals, such as background graphics or redundant tick-marks, are removed (Blasio & Bisantz, 2002; Gillan & Richman, 1994). Maintaining a high *data-ink ratio*, where the highest possible proportion of a visualization's "ink" is devoted to presenting data information (Tufle, 1983), has been shown to improve users' success in interpreting pie charts (Schonlau & Peters, 2008), bar charts (Cleveland & McGill, 1984), and dynamic graphs (Blasio & Bisantz, 2002).

However, others have argued that, when used appropriately, visual embellishment and redundancy can play a supporting role in fostering user understanding by making visualizations more memorable, cohesive, and contextually relevant. Borkin and colleagues demonstrated that contextual detail and human-recognizable images increase user-recall and argue that this increased memorability cannot be separated from understanding (Borkin et al., 2013). Others have argued that redundant coding, for example, making two symbols vary in both shape and color instead of in a single dimension, is a useful way to separate variables and improve comprehension (Egeth & Pachella, 1969; Eriksen & Hake, 1955; Ware, 2012). Bateman and colleagues found that charts embellished with thematically relevant imagery improved user recall weeks later, without hampering users' initial ability to interpret the charts' messages (Bateman et al., 2010). Kosara and Mackinlay (2013) point out that visual embellishments traditionally classified as "chart junk" can be an effective means of bringing the power of storytelling to data visualizations by providing context and increasing memorability. And others have pointed out that visual embellishments can influence users' arousal and emotional responses, calling for more comprehensive measures of visualization-effectiveness than accuracy and response time alone (Tractinsky & Meyer, 1999).

The role of embellishment in the visualization of large data sets remains a controversial subject. We contribute to this ongoing debate by examining the role of visual detail in an interactive visualization built on metaphor. Given the effectiveness of visual details in improving memorability, we hypothesize that providing metaphorical context through visual embellishment will improve users' recall of its functionality.

Landscape and Spatialization Metaphors

We use a landscape metaphor in our visualization (described in Materials and Methods) to display a network of academic fields. This is a common technique for presenting stores of knowledge or information visually, also known as the "distance-similarity metaphor" (Montello, Fabrikant, Ruocco, & Middleton, 2003). Prior work has, for example, plotted academic papers clustered by discipline in 2D space (Woon & Madnick, 2012), presented document themes as topographical features (Granitzer, Kienreich, Sabol, Andrews,

[1]We follow Lakoff and Johnson's convention of small caps for image schemas (Lakoff, 1993).

DOI: 10.1002/asi

& Klieber, 2004), and generated map-like representations of Wikipedia articles complete with bodies of water and land masses (Pang & Biuk-aghai, 2011). We argue that the recurrence of this metaphor in visualizations of knowledge is not a coincidence but stems from the fact that it aligns with an existing image schema.

Prior work has laid out theoretical framing justifying the use of spatial image schemas for the presentation of information (Fabrikant & Buttenfield, 2001; Risch, 2008). These articles explain that cartographers have used spatialization metaphors to visualize information for thousands of years and provide a systematic understanding of why this approach is effective. Prior work has also empirically evaluated the usability of a tool for a large-scale visualization built on a spatialization metaphor and intended to foster the perception that physical closeness is representative of data similarity (Fabrikant & Buttenfield, 2001). We build on this prior work by looking beyond spatial schemas and examining the known set of schemas in its entirety when creating our visualization tool. Further, we explore the interaction between schema-alignment and visual embellishment.

Materials and Methods

Based on this background literature, we used known image schemas to guide us in creating a visually embellished, interactive visualization[2] of a large data set. Specifically, the tool displays the fields and subfields of academia and the connections between them, using the citation network of all 2 million JSTOR articles as the underlying data set. We created four different versions of the tool to manipulate i) whether or not data were presented in a way that was congruent with known image schemas, and ii) whether or not the tool was embellished with visual details to evoke a metaphor.

Data Set

We used the JSTOR library, a digital archive of 2 million academic papers connected by more than 8 million citations and dating back to 1665, as the underlying data set for our visualization. We extracted the full text and bibliographic data from all papers and constructed a citation graph where the nodes represent papers and the links represent citations.

This network is hierarchical (e.g., the field of biology is composed of the subfields cellular biology, field ecology, etc.), and one can infer the domains and subdomains from the patterns and density of citations in this graph (e.g., biology papers are more likely to cite biology papers than papers in the humanities). There are many methods for inferring these clusters (Rosvall & Bergstrom, 2011), and we used the well-known network clustering algorithm, InfoMap, for this purpose. InfoMap specifies releasing a random walker on a network and recording its steps; the most efficient description of the random walk reveals the major patterns (or clusters) of the network. For a more detailed description, see Rosvall, Axelsson, and Bergstrom (2009).

This clustering resulted in the identification of 1,765 fields and subfields and the links among them. We used InfoMap to establish the communities (Rosvall & Bergstrom, 2011; West, Wesley-Smith, & Bergstrom, 2016) and then examined the top 50 papers in each community and made a subjective assessment as to the field label. The labels were first established for a previous paper using the JSTOR corpus (West et al., 2013). Two of the authors of that study spent many hours going through each of the 1,765 fields. The relatedness, or distance between fields, was determined using the average shortest citation paths between all papers in one field and all papers of another field (Newman, 2001). The influence of one field relative to another was determined using the article-level Eigenfactor score to rank and map article-level citation networks (Rosvall & Bergstrom, 2011). Thus, our final data set contained the extracted fields, their hierarchical structure, a "relatedness" score for each pair of fields (measured by citations), the number of papers in each field, and the total influence of each field. We then set out to create an interactive visualization to display these properties to the user.

Design Process

Mapping data attributes to visual variables. To construct an interactive visualization to present this network, we first used existing image schemas to map each data attribute to a specific visual variable. For each academic field, we presented: strength of the relationship to each related field, total number of papers per field, and influence of a field. In order to determine how to represent each data attribute visually, we consulted Lakoff's master metaphor list (Lakoff, 1994) and the Image Schema Catalogue (ISCAT), a database of known schemas (ISCAT, n.d.). Each of these resources lists established image schemas along with descriptions and examples from everyday language.

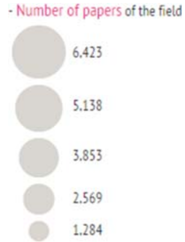
We reviewed these catalogs in their entirety to identify an image schema that described each data attribute. We identified SIMILARITY IS CLOSENESS as the most relevant schema to represent the relatedness of two fields. We identified MORE IS BIGGER as the most relevant schema for representing the number of papers in each field, and IMPORTANCE IS WEIGHT as a guiding principle for presenting a field's influence.

We next matched an established visual variable to each of these schemas. To visually capture the SIMILARITY IS CLOSENESS schema, we represented relatedness between fields using the visual variable of distance, with more-related fields closer together and less-related fields further apart. To visually capture the MORE IS BIGGER schema, we represented the number of published papers in a field with the visual variable size, with a greater number papers represented with a larger circle and a smaller number of papers represented with a smaller circle. Finally, we attempted to visually capture the INFLUENCE IS WEIGHT

[2]All four interactive versions of the tool can be found at: http://efvis.mooo.com/

DOI: 10.1002/asi

TABLE 1. Mapping between data attributes and visual variables. [Color table can be viewed at wileyonlinelibrary.com]

Data attribute	Relatedness between fields	Number of published papers	Degree of influence
Schema	SIMILARITY IS CLOSENESS	MORE IS BIGGER	IMPORTANCE IS WEIGHT
Visual Variable	Position (2D distance between data points)	Size (Area of data point)	Color Saturation
Excerpt from Screenshot	 *(screenshot of map interface)*	 *(screenshot of legend)*	 *(screenshot of legend)*

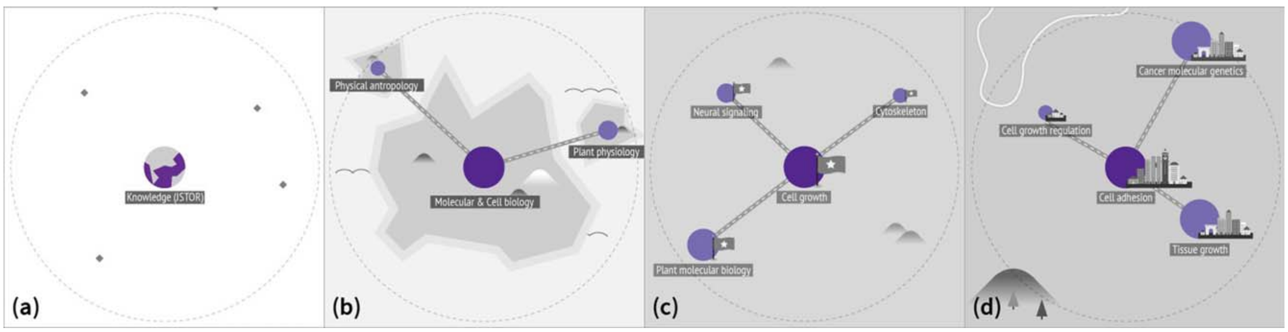


FIG. 1. The first four levels (of a possible seven) of a metaphorical visualization of hierarchical data set representing the fields of academia. As the user zooms in from world (a), to continent (b), to country (c), to city (d), s/he progressively enters a more specialized subfield, descending here from **m**olecular & cell biology to cell growth to cell adhesion. [Color figure can be viewed at wileyonlinelibrary.com]

schema, which we found challenging, as "weight" is not typically considered a visual property. Because prior work shows that humans perceive more-saturated colors to be physically heavier and less-saturated colors to be physically lighter (Alexander & Shansky, 1976), we represented the influence of the field with color saturation. Table 1 shows the mapping between each data attribute and: i) its associated schema, ii) its associated visual variable, and iii) the resulting graphical output.

Adding visual embellishment. To determine if there was an appropriate high-level metaphor for our citation-network data set, we followed the process used to generate the existing catalog of known image schemas (Lakoff, 1994). We conducted a series of brainstorming sessions to identify commonplace expressions that discuss academia in metaphorical language. We identified an abundance of phrases that describe academia as a landscape, such as, "Her *area* of interest is astrophysics," "His work is at the *intersection* of social media and crisis management," "He is interested in the *field* of Economics," or "Her research *explores* online collaboration."

Much like the fact that the abundance of phrases which describe "argument" as "war" reflect an underlying, widespread understanding that argument and war share analogous structures, these statements reflect an existing metaphorical understanding that academic disciplines and the ways in which they are linked share structural features with physical space. As a result, we decorated our tool with metaphor-appropriate visual embellishments, such as roads, mountains, and cities.

Visualization Tool

The result of this design process was an interactive map-like interface that supports zooming and panning over academic fields. As a user zooms in on a particular field, s/he can descend into its subfields, which are connected by roads. Grayscale background imagery displays details of a physical landscape, including landmasses, water features, and buildings. As the user zooms deeper into the visualization, these assets reflect an increasingly localized landscape (progressing, for example, from a view of multiple continents into a street-level view). Figure 1 shows screenshots taken at four different zoom levels.

Across this landscape, each field is represented by a purple dot that reflects the field's total number of papers (size) and total influence on all related fields (color saturation). The field's degree of citation-relatedness to the current field-of-interest was mapped to physical distance from the field-of-interest, and we used the D3 force-directed layout algorithm to convert the collective set of distances into

DOI: 10.1002/asi

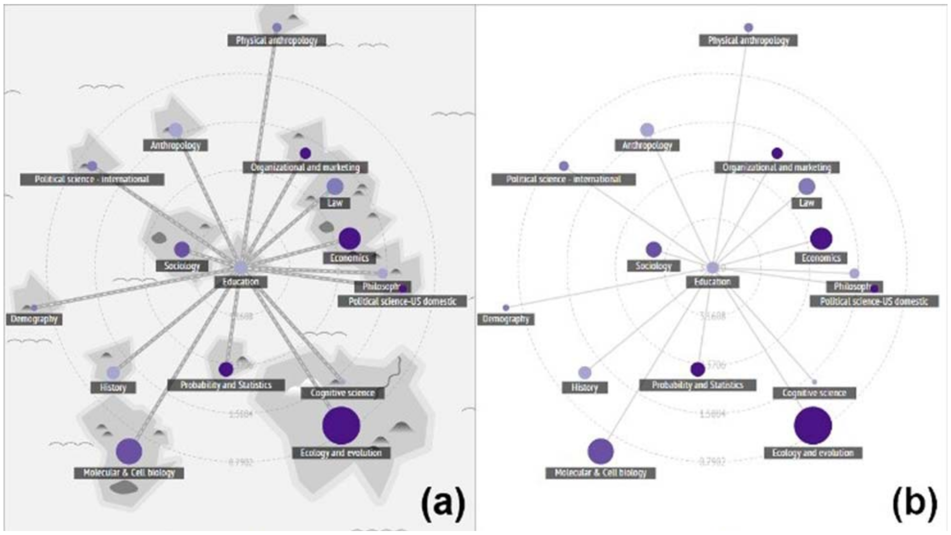


FIG. 2. (a) Embellished and (b) abstract versions of the tool displaying the same data. [Color figure can be viewed at wileyonlinelibrary.com]

physical positions. We were constrained by the fact that relationships among fields could not all be presented simultaneously in 2D space and were forced to reperform this layout when the user selected a new field-of-interest.

The tool presents the user with several mechanisms for interacting with the content. In addition to zooming in and out to see parent and child fields, the user can select any field to set it as the current "field of interest." This positions the selected field at the center of the map and arranges each related field such that its distance from the field of interest reflects the strength of the relationship between the two.

System Variants

We wanted to explore the contribution of both i) schema-congruency, and ii) visual embellishment to users' understanding of the data and experience using the tool. Thus, we independently altered two elements of the interface, its schema alignment and its visual embellishment, to create a total of four different versions of the tool.

Removing schema congruency. To create schema-incongruent versions of the tool, we remapped our data attributes to new visual variables. We took the same three visual variables that we used for our schema-congruent version of the tool (position, size, and color saturation), and assigned a random number to each. We used this random ordering to remap the same three data attributes (relatedness, number of papers, and influence) to new variables. In this way, we hoped to isolate the effect of schema-congruency on users' experiences. In the remapped, schema-incongruent versions of the tool, relatedness was represented by color saturation, number of papers was represented by position, and influence was represented by size.

Removing embellishment. In all four versions of the tool, each academic field was represented by a dot. In the embellished version of the tool, these dots were spread over background visuals of a physical landscape. To remove embellishment, we simply removed all background assets, leaving only the data ink (the dots representing fields) and chart elements (isograms marking distance and the legend). The interface with and without embellishment is shown in Figure 2.

Thus, we created four implementations of the system:

1. **Congruent-Embellished**: Schema-congruent mappings between data attributes and visual variables and background images containing landscape visuals
2. **Congruent-Minimal**: Schema-congruent mappings between data attributes and visual variables with no background images
3. **Incongruent-Embellished**: Data attributes were remapped onto their alternative, schema-agnostic visual variables. Background images contained landscape visuals
4. **Incongruent-Minimal Condition**: Data attributes were remapped onto the same schema-agnostic visual variables. Background images were removed

Assessment Materials

Each participant was randomly assigned to work with exactly one version of the tool. Participants viewed their assigned tool within a surrounding assessment portal. This shell was identical across conditions. The shell sequentially presented the user with a series of tasks (identical for all conditions):

Tutorial and open exploration. The assessment portal first presented a tutorial that walks the user through the main features of the tool. Once the user has completed the tutorial, the portal asked the user to experiment with the tool.

Look-up tasks. After 2 minutes of free exploration, the portal asks a series of seven look-up questions which require the user to interact with the tool and then select a response; for example, "Which field is most closely related to the field of Law?" The first six tasks were short-answer or forced-choice (see Appendix) and covered all of the data attributes we presented (relatedness, size, influence). Participants were instructed to answer each question as quickly as they could without sacrificing accuracy.

Tool comprehension tasks. After the user completed all look-up tasks, the interactive map was temporarily disabled, and the assessment portal then presented nine Likert-style questions to assess how well the user understands and recalls the features of the tool (see Appendix). For example, one question asks the user to select the extent to which he or she agrees with the statement "This tool can be used to determine whether two fields are related."

Open-ended data exploration tasks. After the user completed all nine tool-comprehension questions, the visualization again became visible and interactive. The assessment then prompted the user to answer three open-ended questions (see Appendix); for example, "Imagine you are a researcher in the field of Anthropology. What kinds of collaborators would you reach out to and why?" allowing her or him to interact with the tool as needed. These questions were intended to provide an opportunity for the user to draw inferences from the data presented and consider the data presented in a real-world context. Finally, we collected self-reported enjoyment, understanding, and demographic information.

Participants

We conducted our evaluation online with adult participants. We recruited through e-mail solicitations posted to distribution lists at multiple universities. Because the visualization presents a view of academia, we advertised to groups with academic ties in an attempt to increase the likelihood that participants would have an interest in the underlying data set. A total of 145 adults enrolled in the study. Thirty participants evaluated a fifth version of the tool whose design is outside the scope of this paper.

Thirty-five participants dropped out before completing all tasks and questions. We collected complete data on the four versions of the tool analyzed here from 80 adults (33 women, mean age 27, SD = 3.7). As we continued recruiting to replace dropouts until 20 participants had completed all tasks in each of our four conditions (listed below), participants were spread evenly across conditions with 20 individuals in each group.

Procedures

Participants volunteered for the study by navigating directly to the online assessment portal via a recruitment e-mail. The assessment portal guided the participant through each section of the task as described above: tutorial, free exploration, look-up tasks, tool-comprehension tasks, open-ended tasks, and collection of enjoyment and demographic information. Across conditions, average total task-completion time was 30 minutes (SD = 12 minutes). All participants received a $10 gift card to Amazon as a thank you for their participation.

Data Analysis

We recorded the total number of questions completed for all participants, including those who dropped out. All other measures were calculated only for the 80 participants who completed all required tasks.

We recorded response time (in seconds), accuracy (0 or 1), and total number of mouse-clicks for each of the short-answer and forced-choice look-up tasks. For the tool comprehension questions, we measured correctness on a scale of 1 to 5, with *strongly agree* responses receiving the highest correctness score for true statements and *strongly disagree* responses receiving the highest correctness score for false statements. For each participant, correctness scores were averaged across all nine questions for a single correctness score.

Comments from the three open-ended questions were coded by a researcher who was blind to participants' assigned conditions. We labeled a response as "factual" if it reported straightforward information displayed by the tool without interpretation, such as: "There are 16 subfields in the field of economics." We labeled a response as "interpretive" if the user interpreted the information and drew conclusions not immediately apparent, or added his or her own feelings, thoughts, or opinions. Through iterative coding, we developed the following list of criteria for interpretive responses: i) the participant drew conclusions by combining information from the tool and their personal experience; ii) the participant pointed out unexpected findings revealed by the tool; iii) the participant posed new questions in response to the information presented in the tool; and iv) the participant made recommendations about how the tool might foster additional ways of engaging with the data. A second researcher, also blind to condition, coded a randomly selected 50% of all responses to open-ended questions. Cohen's κ was 0.974 ($p < .001$).

Results

Completion Rate

We collected progress data on each participant who began our study and measured the total number of tasks completed, regardless of whether the participant dropped out before finishing (those who dropped out were not included in our final data set, and their responses are not included in subsequent analyses). Of the 115 individuals who began the study, 35 dropped out early. The fraction of participants who did not complete all tasks is shown in Figure 3 by condition.

The distribution of the number of questions completed was highly bimodal (i.e., participants were most likely to drop out early or to persist to the end); thus, we used

DOI: 10.1002/asi

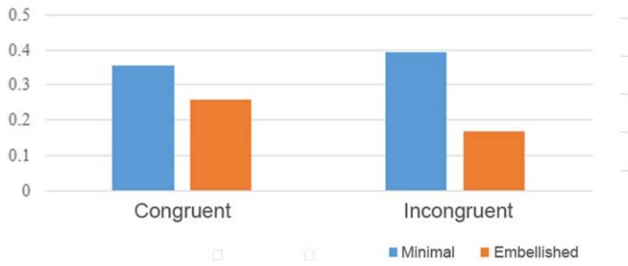


FIG. 3. Fraction of participants who dropped out of the study before completing all tasks, by condition. [Color figure can be viewed at wileyonlinelibrary.com]

nonparametric statistical tests to compare completion-rate distributions across groups. A Mann–Whitney U-test revealed that the participants working with embellished versions of the tool ($N_1 = 54$) completed significantly more tasks than participants working with minimal versions of the tool ($N_2 = 65$, $U(1) = 4.0$, $p = .046$). Thus, participants were more persistent when working with a version of the tool embellished with contextual landscape detail. There was no significant effect of level of congruency.

Look-up Task Performance

Between conditions, we compared: i) average response time on correct trials, ii) average accuracy, and iii) number of clicks. Because none of these measures were normally distributed, we compared group performance using nonparametric tests. A Kruskal–Wallis test revealed no significant effect of condition on response time. We conducted Mann–Whitney U-tests to examine main effects of level of congruency and level of embellishment. This revealed that participants in congruent conditions were significantly faster (median [Mdn] = 35.05, interquartile range [IQR] = 13.00) than participants in incongruent conditions (Mdn = 39.63, IQR = 20.98, $U = 545$, $p = .021$) (see Figure 4). We removed one outlier who spent several minutes on each task (more than twice as long as the next-slowest participant); however, the pattern of significance remained unchanged regardless of whether this participant was included or excluded. A Mann–Whitney U-test revealed no significant main effect of level of embellishment on response time.

There were no significant differences between groups on average accuracy. Thus, users working with congruent versions of the tool were able perform information-seeking tasks more quickly than those working with incongruent versions of the tool without sacrificing correctness.

Finally, we examined each participant's click-count across look-up tasks as a proxy for understanding the complexity of executing these tasks in each version of the tool. Because the presentation of the hierarchical structure of the data remained the same across conditions, the same information could be retrieved from each tool with an equivalent number of clicks if the user behaved with perfect efficiency. Participants made significantly more clicks when using incongruent versions of the tool (Mdn = 18.5, IQR = 3.75)

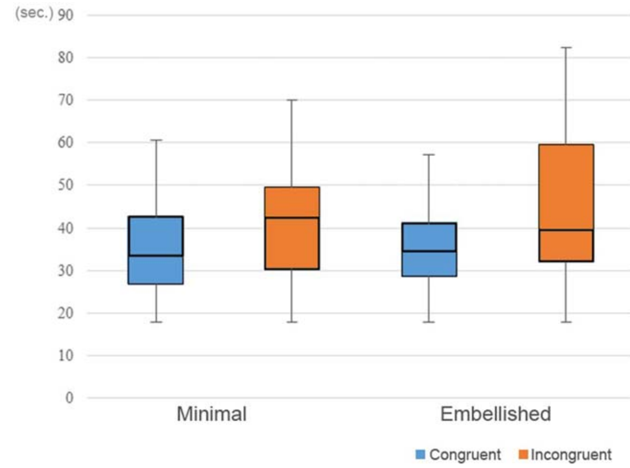


FIG. 4. Average response time by condition on correct responses to look-up questions. [Color figure can be viewed at wileyonlinelibrary.com]

than when using congruent versions of the tool (Mdn = 16, IQR = 2.75, $U = 515$, $p = .006$). There was no significant effect of level of embellishment on click-efficiency. Thus, users in incongruent conditions performed more steps than participants in congruent conditions to retrieve the same information.

Tool Comprehension

We next compared participants' accuracy on tool-comprehension questions. We used parametric statistical tests, as scores were approximately normally distributed. A two-way analysis of variance (ANOVA) of the effect of level of congruency and level of embellishment on average correctness revealed a significant interaction between these two independent measures ($F(1, 76) = 6.232$, $p = .015$, $\eta^2 = .076$).

We conducted a posthoc analysis of simple main effects. Users who worked with congruent versions of the visualization had significantly higher correctness scores when they saw embellished visuals (3.918 ± 0.20) than when they saw minimal ones. (3.582 ± 0.20, $F(1, 76) = 5.579$, $p = .021$, $\eta^2 = .068$). Among users who worked with incongruent versions of the visualization, there was no significant difference on average correctness between those who saw embellished visuals (3.633 ± 0.20) and those who saw minimal ones (3.799 ± 0.20).

Users who worked with embellished versions of the tool had significantly higher correctness scores when they saw the congruent version ($F(1, 76) = 4.026$, $p = .048$). There was no significant difference between the congruent and incongruent groups who saw the minimal version of the tool. Thus, for our users, embellishment played a role in facilitating comprehension only when the underlying structure of the metaphor was congruent with users' image schemas. When users worked with incongruent visual metaphors, embellishment had no benefit.

Frequency of Data Inferences

We coded each response to each open-ended question as either "factual" or "inferential." For example, in response to the prompt: "Now navigate until you are looking at the field of Economics. Please describe an interesting thing about this field in the box below," participants' factual statements included:

> Classical studies is not highly related nor very influential to the field of Economics.
>
> 128,131 papers published in 2013. Influence of the field was 8.959 Eigen factor. There are 16 subfields of Economics. "Econometric analysis" is the most popular subfield.
>
> The most influential papers are math related area.

Inferential statements included:

> It does not have the most papers published but however it is highly influential to knowledge, meaning that the importance of knowledge has nothing to do with how many paper is published.
>
> All of the related fields have a high number of publications. (Not sure how to interpret this – perhaps something about "Economics" being too broad of a topic to really infer relationships from?)
>
> Operations in research is very related, but with few publications. Must be a very specialized field?

For each participant, we coded whether s/he made any inferential statements in response to any of the open-ended prompts. We then examined the effect of level of embellishment and level of congruency on the presence of inferences. We conducted two nominal chi-square tests, one for association between level of embellishment and presence of inferential statements, and one for association between level of embellishment and presence of inferential statements. For both tests, all expected cell frequencies were greater than 5. Because we did not have a hypothesis about whether either factor would influence users' frequency of making inferential statements, we applied a Bonferroni adjustment to all p-values. The first test revealed that participants working with embellished versions of the tool were significantly more likely to make an inferential comment than users working with minimal versions of the tool ($\chi^2(1) = 5.051$, $p = .025$). The second test revealed that there was no significant difference in frequency of inferential statements based on level of congruency.

Discussion

Effects of Schema-Congruency

Our first hypothesis, that users will be more efficient when working with metaphorical visualizations that align with documented image schemas, was supported by our results. Participants spent less time retrieving identical information from schema-congruent versions of our tool than from schema-incongruent ones, and look-up tasks required fewer clicks in schema-congruent conditions. These efficiency gains are consistent with the idea that visualizations aligned with known image schemas can support cognitive offloading by leveraging users' preexisting structural understanding of the world.

Our schema-congruent and schema-incongruent tools were identical except that the mappings between data attributes and visual variables were randomly reassigned in the schema-incongruent version. Thus, all data attributes, visual variables, patterns of interaction, and numbers of steps to retrieve specific information remained the same across versions. That users had significantly different experiences looking up data values despite the uniformity across interfaces is consistent with the idea that manipulating mappings to align with users' image schemas provides a valuable increase in user efficiency. Prior work provides many suggestions for mapping data to visual variables (Bertin, 1983; Mackinlay, 1986); we contribute to this body of work by providing an additional mechanism for guiding such decisions. As the complexity of a data set grows and data attributes become more numerous, mapping attributes to visual variables becomes more challenging and more likely to be arbitrary. Consulting known image schemas for design guidance promises to expand the tools at a designer's disposal for making these mappings as intuitive as possible.

Effect of Visual Embellishment

Independent of schema-congruency, our results show that adding visual embellishment to our interactive tool increased users' likelihood of completing the study and also led to more inferential statements about the underlying data. One interpretation consistent with both of these findings is that embellishment increased engagement with the tool, relative to the abstract graphical treatment of the minimal version of our tool.

Future work remains to understand if any visual embellishment would have this effect, or if we saw these results because we used the visual details of ACADEMIA IS LANDSCAPE, a schema reflected in everyday language on this topic. Would users be equally engaged with visual detail showing academia as a tool bench? As a dinner plate? As a construction site? Future work remains to understand whether adding visual detail from an arbitrary metaphor is experienced by users as bizarre and disengaging or whether users can be drawn in with embellishment of any kind. Our results show that, at a minimum, embellishment that is derived from a known schema predicts that users will proactively engage with the data, as reflected by our users' increased willingness to continue performing tasks and increased proactive extrapolation and interpretation of the data points they saw.

Combining Schema-Congruency and Embellishment

Most notably, the greatest gains in our participants' understanding came from adding metaphor-appropriate

DOI: 10.1002/asi

embellishment on top of schema-congruent mappings. On our recall test, users were most successful in accurately describing the purpose, features, and capabilities of the tool if they had seen the version that was both schema-congruent and embellished. We hypothesized that embellishment would lead to increases in understanding by giving users a better picture of the overall tool, and thus making it more memorable and cohesive. Our results show that this is the case if, and only if, the underlying tool is aligned with users' internalized understanding of the world.

This suggests that in the design of metaphorical interfaces, the question of embellishment may be *when* it is appropriate rather than *whether* it is appropriate at all. For our participants, recalling the purpose of our metaphorical tool was easiest when they had been exposed to a metaphor that was captured both with structure and with surface detail. Prior work shows that the power of metaphor comes from the fact that it aligns underlying structures between unrelated concepts (Gentner, 1983), thus we might suspect that focusing on surface detail by adding visual embellishment would distract users from accessing the deeper structure of the metaphor. However, our results are instead consistent with the idea that visual detail acts as a means of communicating the domain of the metaphor. Thus, designers creating metaphorical visualizations for large data sets may do well to embellish them with metaphorical surface detail—but only if they have first created meaningful underlying structure.

Limitations and Future Work

The design choices that we were led to by following the guidance of existing image schemas have much in common with the guidance we would have received if we had relied on existing design principles. One challenge of evaluating image schemas as a design tool is that they are, by definition, intuitive. As a result, there is a great deal of overlap between these schemas and what we might consider "good design." This can be seen as a potential confound of our results, although it is also the reason that image schemas have so much to offer designers. Future work remains to repeat this process with other data sets and across diverse contexts.

Another limitation of this work is the false dichotomy of describing a visualization as strictly "schema-congruent" or "schema-incongruent," and similarly, as strictly "embellished" or strictly "minimal." In reality, each of these axes is a continuum of possibilities and designers have and will continue to create visualizations at all points in between. Future work remains to understand the effects of incremental changes in each of these properties on user understanding and experience.

Further work also remains to evaluate possible alternate hypotheses that cannot all be tested within the scope of a single study. Expanding this approach to less conventional metaphors and to designs that combine existing design principles with guidance from schemas would help elucidate the extent to which these effects are generalizable and to further isolate the contribution of metaphor to the outcome measures we examined.

Conclusion

Our results show that image schemas can serve as valuable guideposts for creating interactive visualizations for large data sets. Whether an interface is more intuitive when rendered as a flowing stream or as a set of interconnected roadways may well depend on the schemas users employ when thinking about it. These metaphorical structures are already present in everyday thought and speech, ready to provide powerful and easily accessible guidance to designers.

Our results indicate that users may find these metaphorical interfaces more engaging and easier to understand if they include surface detail evoking the underlying schema-aligned metaphor. Our results suggest that visual details of a metaphor can draw users in and hold their attention. We saw evidence that when these details are reflective of a deeper structure that is an appropriate representation of the structure of the data set, they improve user understanding and cement the tool into a cohesive whole.

Both previously identified schemas and novel ones extracted from everyday language can illuminate metaphors that are appropriate for a particular data set or particular data attribute. Designers can bring these metaphors to life with visual detail evoking the context of the metaphor's domain, thereby increasing users' awareness of the metaphor and engagement with the visualization. We have long known that in visualizations, metaphor can be a workhorse, a guiding light, a shroud, or a double-edged sword. Our results show predictive ways in which specific design choices lead to these effects and leave us a step closer to understanding why.

Acknowledgments

We thank JSTOR for providing the bibliographic data for building the map of science.

References

Alexander, K.R., & Shansky, M.S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors. Perception & Psychophysics, 19, 72–74.

Barr, P., Khaled, R., Noble, J., & Biddle, R. (2005). A taxonomic analysis of user-interface metaphors in the Microsoft Office Project Gallery. Proceedings of the Sixth Australasian Conference on User Interface (pp. 109–117). Retrieved from http://crpit.com/confpapers/CRPITV40Barr.pdf%5Cnhttp://dl.acm.org/citation.cfm?id=1082257

Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., Mcdine, D., & Brooks, C. (2010). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In Human factors, (pp. 2573–2582).

Bertin, J. (1983). Semiology of Graphics. Physical Review Letters (Vol. 94). Retrieved from http://portal.acm.org/citation.cfm?id=1095597

Blackwell, A.F. (2006). The reification of metaphor as a design tool. ACM Transactions on Computer-Human Interaction, 13, 490–530.

DOI: 10.1002/asi

Blackwell, A.F., & Green, T.R.G. (1999). Does metaphor increase visual language usability? In Proceedings 1999 IEEE Symposium on Visual Languages.

Blanchette, I., & Dunbar, K. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. Memory & Cognition, 29, 730–735.

Blasio, A.J., & Bisantz, A.M. (2002). A comparison of the effects of data-ink ratio on performance with dynamic displays in a monitoring task. International Journal of Industrial Ergonomics, 30, 89–101.

Boers, F. (1999). When a bodily source domain becomes prominent: The joy of counting metaphors in the socio-economic domain. Amsterdam Studies in the Theory and History of Linguistic Science Series, 4, 47–56.

Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable. IEEE Transactions on Visualization and Computer Graphics, 19, 2306–2315.

Chau, M. (2011). Visualizing web search results using glyphs: Design and evaluation of a flower metaphor. ACM Transactions on Management Information Systems, 2 (1), 2.

Cleveland, W.S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association, 79, 531–554.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), The nature of insight (pp. 365–395). New York: MIT Press.

Egeth, H., & Pachella, R. (1969). Multidimensional stimulus identification. Perception & Psychophysics, 5, 341–346.

Eriksen, C.W., & Hake, H.W. (1955). Absolute judgments as a function of stimulus range and number of stimulus and response categories. Journal of Experimental Psychology, 49, 323.

Fabrikant, S.I., & Buttenfield, B.P. (2001). Formalizing semantic spaces for information access. Annals of the Association of American Geographers, 91, 263–280.

Fischer, S., Itoh, M., & Inagaki, T. (2014). Prior schemata transfer as an account for assessing the intuitive use of new technology. Applied Ergonomics, 46, 8–20.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy*. Cognitive Science, 7, 155–170.

Gentner, D., & Nielsen, J. (1996). The Anti-Mac interface. Communications of the ACM*, 39*, 70–82.

Gillan, D.J., & Richman, E.H. (1994). Minimalism and the syntax of graphs. Human Factors: The Journal of the Human Factors and Ergonomics Society, 36, 619–644. Retrieved from http://hfs.sagepub.com/content/36/4/619.abstract

Gordon, W.J.J. (1974). Some source material in discovery-by-analogy. The Journal of Creative Behavior, 8, 239–257.

Graham, H., Yang, H., Berrigan, R. (2004). A solar system metaphor for 3D visualisation of object oriented software metrics. In Proceedings of the 2004 Australasian Symposium on Information Visualisation-Volume *35* (pp. 53–59). Retrieved from http://dl.acm.org/citation.cfm?id=1082108

Granitzer, M., Kienreich, W., Sabol, V., Andrews, K., & Klieber, W. (2004). Evaluating a system for interactive exploration of large, hierarchically structured document repositories. In IEEE Symposium on Information Visualization.

Hamilton, A. (2000a). Interface metaphors and logical analogues: A question of terminology. Journal of the American Society for Information Science and Technology, 51, 111–122.

Hamilton, A. (2000b). Metaphor in theory and practice: The influence of metaphors on expectations. ACM Journal of Computer Documentation (JCD), 24, 237–253.

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting InfoVis with visual difficulties. IEEE Transactions on Visualization and Computer Graphics, 17, 2213–2222.

Huron, S., Vuillemot, R., & Fekete, J.D. (2013). Visual sedimentation. IEEE Transactions on Visualization and Computer Graphics, 19, 2446–2455.

Hurtienne, J., & Blessing, L. (2007). Design for intuitive use — Testing image schema theory for user interface design. In ICED 07 Paris, 16th International Conference on Engineering Design, Proceedings of the conference (pp. 1–12).

ISCAT. (n.d.). Retrieved from http://iscat.stefciu.de/. Accessed January 1, 2016.

Kosara, R., & MacKinlay, J. (2013). Storytelling: The next step for visualization. Computer, 46, 44–50.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), Metaphor and thought (pp. 202–251). Cambridge, MA: Cambridge University Press.

Lakoff, G. (1994). Master metaphor list. Berkeley, CA: University of California.

Lakoff, G., & Johnson, M. (1997). Metaphors we live by. In J. O'Brien & P. Kollock (Eds.), The production of reality: Essays and reading on social interaction (pp. 124–134). Chicago: University of Chicago Press.

Loeffler, D., Hess, A., Maier, A., Hurtienne, J., & Schmitt, H. (2013). Developing intuitive user interfaces by integrating users' mental models into requirements engineering. In Proceedings of the 27th International BCS Human Computer Interaction Conference (p. 15:1–15:10). Swinton, UK: British Computer Society. Retrieved from http://dl.acm.org/citation.cfm?id=2578048.2578069

Löffler, D., Lindner, K., & Hurtienne, J. (2014). Mixing languages': Image schema inspired designs for rural Africa. In CHI'14 Extended Abstracts on Human Factors in Computing Systems (pp. 1999–2004).

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. ACM Transactions on Graphics, 5, 110–141.

Montello, D.R., Fabrikant, S.I., Ruocco, M., & Middleton, R.S. (2003). Testing the first law of cognitive geography on point-display spatializations (pp. 316–331). Berlin: Springer.

Newman, M.E.J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E, 64(1Pt2), 16132.

Pang, C., & Biuk-Aghai, R.P. (2011). Wikipedia world map: Method and application of map-like Wiki visualization Wiki category pre-processing. In *WikiSym '11* (pp. 124–133). Mountain View, CA.

Risch, J.S. (2008). On the role of metaphor in information visualization, 20. Human-Computer Interaction; Graphics. Retrieved from http://arxiv.org/abs/0809.0884

Rosvall, M., Axelsson, D., & Bergstrom, C.T. (2009). The map equation. European Physical Journal: Special Topics, 178, 13–23.

Rosvall, M., & Bergstrom, C.T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PLoS One, 6, e18209.

Santos, C.R., Dos, Gros, P., Abel, P., Loisel, D., Trichaud, N., & Paris, J.P. (2000). Mapping information onto 3D virtual worlds. In 2000 IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics.

Schonlau, M., & Peters, E. (2008). Graph comprehension: An experiment in displaying data as bar charts, pie charts and tables with and without the gratuitous 3rd dimension. Social Science Research Network Working Paper Series. Retrieved from http://ssrn.com/abstract=1272627

Thomas, J.J., & Cook, K.A. (2006). A visual analytics agenda. IEEE Computer Graphics and Applications, 26, 10–13.

Tractinsky, N., & Meyer, J. (1999). Chartjunk or goldgraph? Effects of presentation objectives and content desirability on information presentation. MIS Quarterly, 23, 397–421.

Tufle, E. (1983). The visual display of quantitative information. *CT Graphics, Cheshire*. Retrieved from http://www.colorado.edu/UCB/AcademicAffairs/ArtsSciences/geography/foote/maps/assign/reading/TufteCoversheet.pdf

Ware, C. (2012). Information visualization: Perception for design. Amsterdam: Elsevier.

West, J.D., Jacquet, J., King, M.M., Correll, S.J., Bergstrom, C.T., Moss-Racusin, C. . . . Paik, I. (2013). The role of gender in scholarly

DOI: 10.1002/asi

authorship. PLoS One, 8, e66212. http://doi.org/10.1371/journal.pone.0066212

West, J.D., Wesley-Smith, I., & Bergstrom, C.T. (2016). A recommendation system based on hierarchical clustering of an article-level citation network. IEEE Transactions on Big Data, 2, 113–123.

Woon, W.L., & Madnick, S. (2012). Semantic distances for technology landscape visualization. Journal of Intelligent Information Systems, 39, 29–58.

Ziemkiewicz, C., & Kosara, R. (2008). The shaping of information by visual metaphors. IEEE Transactions on Visualization and Computer Graphics, 14, 1269–1276.

Appendix: Task Prompts

Short-answer and forced-choice look-up tasks: Participants were asked to perform the following active tasks. Participants were able to use the visualization tool to help them answer these questions.

- Which field is the most related to the field of Law?
- Name one subfield inside Sociology.
- Name one subfield of Economics that is highly influential.
- What is the least influential field related to Bats (a subfield of Ecology and Evolution)?
- Which field related to Organizations and Institutions (a subfield of Organizational and Marketing") has published the most papers?
- Which pair of fields are more related?
  a. Economics and Education
  b. Economics and Sociology
  c. I don't know

Tool comprehension tasks: Participants were asked to evaluate a series of statements about the tool in response to the prompt, "Please state the extent to which you agree with the following statement." Options ranged from 1 – "Strongly Agree" to 5 – "Strongly Disagree," with an additional option outside the Likert scale stating, "I don't understand the tool well enough to answer this question." When answering these questions, participants did not have access to the visualization tool.

- This tool can be used to determine the size of a particular academic field
- This tool can be used to look up the most influential author in a field.
- This tool can be used to determine whether two fields are related.
- This tool can be used to determine when a field became established.
- This tool can be used to determine how fast a field is growing.
- This tool can be used to determine which subfields are inside the field of Law.
- This tool can be used to determine which of two fields is more influential.
- This tool can be used to understand how a field has changed over time.
- This tool can be used to determine which fields are most related to Economics.
- Open-ended data exploration tasks: Participants were asked to answer free-response questions using the visualization tool.
- Now navigate until you are looking at the field of Economics. Please describe an interesting thing about this field in the box below.
- Imagine you are a researcher in the field of Anthropology. What kinds of collaborators would you reach out to and why?
- Imagine that you are the Dean of the School of Education of a small but growing university and have the opportunity to start a new department. What do you do?

DOI: 10.1002/asi