Improving User Confidence in Concept Maps: Exploring Data Driven Explanations

Pierre Le Bras, David A. Robb, Thomas S. Methven, Stefano Padilla, Mike J. Chantler

School of Mathematical and Computer Sciences, Heriot-Watt University

Edinburgh, Scotland, UK EH14 4AS

{pl196, d.a.robb, t.methven, s.padilla, m.j.chantler}@hw.ac.uk

ABSTRACT

Automated tools are increasingly being used to generate highly engaging concept maps as an aid to strategic planning and other decision-making tasks. Unless stakeholders can understand the principles of the underlying layout process, however, we have found that they lack confidence and are therefore reluctant to use these maps. In this paper, we present a qualitative study exploring the effect on users' confidence of using data-driven explanation mechanisms, by conducting in-depth scenario-based interviews with ten participants. To provide diversity in stimulus and approach we use two explanation mechanisms based on projection and agglomerative layout methods. The themes exposed in our results indicate that the data-driven explanations improved user confidence in several ways, and that process clarity and layout density also affected users' views of the credibility of the concept maps. We discuss how these factors can increase uptake of automated tools and affect user confidence.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

Author Keywords

User Confidence; Concept Map; Data Driven Explanation; Qualitative Study.

INTRODUCTION

Concept maps, a type of visualization that spatially organises ideas by similarity, are often used for planning, decision making and other collaborative activities. Examples include: 1) strategic review of organisations' activities and operations, e.g. after mergers or acquisitions; 2) international benchmarking for research institutions and directorates; and 3) understanding product catalogues and customer segmentations. With the growth in access to, and increased ability to process large document corpora, such maps are becoming more common
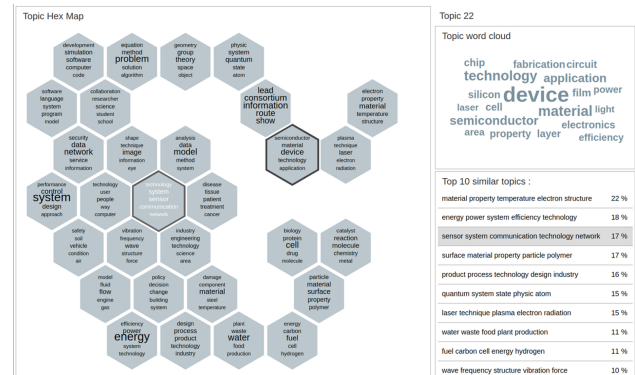
Figure 1. An example of a concept map as presented to our participants. On the left, we show various concepts organized as a concise concept map. On the right, we present additional information about the selected concept (a wordcloud and a list of similar concepts).

and are frequently generated using automated tools [16, 57]. Consequently, stakeholders can now more easily augment their decision making with these informative overviews.

From pilot interviews and discussions with decision makers, planners and management, however, we have found that they are wary of using these automatically generated maps for higher risk activities as they do not feel confident in their ability to explain the layouts to third parties. A particular issue that emerged is the apparent confusion caused by the use of the dimensionality reduction techniques often employed in map generation software. In these situations, while stakeholders appreciated the qualities of similarity-based layouts, they felt that they were not able to defend their decisions, or explain the generation principles of the maps. This was acutely problematic when there were others affected, for example, when they had to report to their supervisors, or when funding was reliant upon their decisions.

In this paper we therefore investigate layout and explanation methods that help users understand the principles behind concept map generation, and in particular which help them improve their confidence in such visualisations. We do this by integrating interactive data-driven explanations with two different forms of concept map generation: the first approach uses standard dimensionality reduction and projection techniques while the second utilises a bottom-up agglomerative method. We call these methods the *reductive* and *constructive* approaches, respectively.

We present a qualitative study that explores the effect on users' confidence of using these two data-driven explanation mechanisms by conducting in-depth scenario-based interviews with ten participants.

It should be emphasised that our objective here is not to perform A/B testing of these two explanation methods, as they are intimately related, and would therefore be confounded, with their layout methods. Instead we are using both the constructive and reductive approaches to generate a rich set of stimuli for use in our qualitative study. A second clarification concerns the generation of the concepts themselves, for which we use Latent Dirichlet Allocation (LDA) [8]. We have chosen to use LDA because it is popular in the literature and we have found it to be effective. We wish to emphasize, however, that this paper focuses on investigating methods for explaining layout algorithms. Automated concept, or topic, generation from document corpora is outwith the scope of this paper.

In summary, the contributions of this paper are:

1. An investigation of the overall effect of the use of interactive data-driven explanations for concept map generation on users' confidence.

2. An analysis of the use of both constructive and reductive approaches for layout algorithms and their impact on the associated explanation methods and consequently users' confidence.

3. A set of four design recommendations (R1 - R4) for the use of automated layout algorithms and their associated data-driven explanation methods.

BACKGROUND AND RELATED WORK

In this section we describe the type of concept map on which we focus, discuss current automated layout algorithms, and situate our work within the area of algorithmic interpretability and user confidence.

Concept Maps for Planning

The purpose of a concept map is to portray the relationships between ideas [16]. There are various styles including hierarchical, sequential/causal, and semantic concept maps [58]. Hierarchical and sequential/causal concept maps, in which the concepts are explicitly connected, have been developed for learning purposes by Novak [39, 41]. These are essentially node link diagrams.

The semantic style of concept map was developed for planning by Trochim [56]. These are arranged based on the similarity of the concepts which is visualized by their proximity. Trochim's participatory process involves the gathering of stakeholder ideas and their formation into a concept map or affinity diagram [25, 27]. This process has been increasingly used for planning and evaluation purposes [57]. Examples of domains in which it has been used include ICT in education [63], mental health [1], chronic disease prevention [3], research agenda production [19], and implicit sharing in collaborative tasks [24]. The layout of these concept maps is often generated automatically [16]. For instance, a similarity matrix is produced from free-grouping stakeholders' ideas, and the layout of the

ideas is determined by a multidimensional scaling (MDS) of the similarity matrix. The ideas are then coalesced into concepts by clustering and these are overlaid on the MDS layout to produce the final map. These concept maps contain elements of human influence, albeit that of a crowd collectively, due to the layout being based on the card sorting of ideas. However, both ideas and similarity data can be sourced through data mining techniques such as topic modelling [7] producing concept maps with no agenda which visualize the concepts in a document corpus e.g. [43].

Of these two types of concept map, the node-link style (hierarchical and sequential/causal) and those using similarity based concept placement (semantic maps), it is the latter, similarity based concept maps on which we focus in this paper.

Automated Layout Algorithms

Although there are good examples of irregular layout algorithms [62, 10, 21], we have found from our pilot interviews with decision makers, that users prefer regular grid layout for readability and aesthetic reasons. We will therefore focus our survey on algorithms which generate regular layouts.

The algorithms for laying out concepts or entities based on similarity data predominantly make use of projection or dimensionality reduction methods such as multi-dimensional scaling (MDS) [37] or Isomap [55]. One exception, incBoard [48], was designed specifically to handle dynamic data and be easily updateable. Their method relies on placing the concepts iteratively based on pairwise similarity.

Of the layout algorithms which use projection methods, some project onto two-dimensions producing irregular layouts whereas others fit their projections to a regular grid. A notable representative of the latter methods is IsoMatch [20], which combines a dimensionality reduction and an assignment algorithm to create a regular layout.

While there is a large class of projection based methods [22, 52, 13] they are reductive in nature, in that they remove information from a complete and complex highly dimensional solution to generate a more intuitive two-dimensional organisation of ideas.

While we have found many examples of these *reductive* layout methods, we could not find examples of space filling layout methods based on more *constructive* approaches i.e. gradually building up complexity from a set of units (for example concepts) by merging them according to their relative similarity information. While there are many well-known agglomerative clustering algorithms which do for visualisations like Dendrograms, to the authors knowledge these have not been converted into a space filling layout approach, as we discuss in this paper.

Interpretability and Confidence

Lack of interpretability has been identified as largely responsible for a reluctance of domain experts to adopt applications of machine learning particularly for safety critical applications [35]. That work set out a landscape of information processing techniques including k-means and principle component analysis (PCA) as examples of clustering and multidimensional

scaling. Machine learning was highlighted as being alone in its inherent lack of interpretability while all other processing techniques were deemed to be ultimately interpretable. Nevertheless a process does not actually need to be uninterpretable (as with machine learning) to be considered a black box [42] and for this to affect whether users accept its results. It is enough that users be unaware of its working [46]. In our experience users not already familiar with dimensionality reduction techniques struggle with the underlying concept and lack confidence in being able to explain results based on them.

Confidence and trust are important factors in users' willingness to accept and use the outputs of a process [47]. There have been a number of works exploring trust in machine learning algorithms e.g. [31, 23, 34, 49], and in particular in recommender systems e.g. [28, 32, 36, 15].

Pieters distinguishes trust from confidence, in that trust is acquired by means of risk assessments, whereas confidence simply reflects a user's impression of system reliability without going into details [43]. Although our explanations "open the black box" of layout mechanisms, we do not consider them as full risk assessments. Making the explanations accessible to users would only give the details of one concept map at a time, helping them to understand its makeup, to see it as reliable, and therefore increase confidence.

In [44], Padilla et al. explored how users construct and change semantic concept maps, while Lim et al. [33] investigated the effects of explanations on user trust and system intelligibility. However, our work is distinct from these two papers, as our primary goal is to investigate the effects of explanations on user confidence in automated concept maps layouts, in particular the interpretation of the layout algorithm itself. To our knowledge there have been no explorations of confidence in relation to visualization explanations prior to this paper.

Summary
There is a substantial body of work on the automated layout of similarity based visualizations including concept maps, most of which use dimensionality reduction. To our knowledge none of the existing concept map layout algorithms are based on agglomerative clustering. As this technique is simple to explain we develop a new layout algorithm based on it. In our experience naive users find dimensionality reduction methods difficult to understand leading to reluctance to use maps based on it for planning and decision making.

Furthermore, this background study as shown little work done regarding data-driven explanation of concept maps layout, particularly their affect on user confidence.

STUDY DESIGN
In this section we explain the design of our study in which we explore the views of users in terms of confidence in concept maps using data-driven explanation mechanisms. To build the explanations we use two examples of concept map layout methods, one reductive and one constructive. In the following subsections, we first set out the aims and research questions. Then we describe the explanation mechanisms developed for the study. Finally, after presenting the pre-study work carried out, we detail the design of the study.

Aim and Research Questions
Our study aimed to investigate how confidently users were able to explain concept maps to others. Specifically, we sought to answer two research questions:

RQ1: What are the overall effects of data-driven explanations on confidence?

RQ2: What are the specific effects of reductive and constructive approaches on confidence?

Design of the Explanation Mechanisms
In order to explore how users can confidently explain concept map layouts, we use data-driven explanations (or DDEs). The main reasoning for creating these DDEs is that they will enable an explanation mechanism that does not require the presence of an expert user guiding naive users. To fully exploit them we formulated three requirements for our DDEs:

- We aim for them to be visual [59], in order to convey as much information as possible, without imposing on users the need to read too much text.

- They should be interactive [26], in order to allow the user to control the explanation. Users should be able to play, pause, fast-forward, back-track or restart the explanation to investigate specific steps in the process presented to them and get feedback on the data provenance at any moment in the explanation.

- Finally, they should be data-driven [9], in order to represent the actual information, making it specific to the particular concept map users are working with at that moment.

In the following subsections, we will present the two explanation mechanisms developed and used in our study:

- A reductive approach, derived from a selected projection mapping method (IsoMatch).

- A constructive mechanism built upon an agglomerative clustering technique and adapted for the layout of concept maps.

We used two different approaches in our study to offer increased diversity of stimuli to our participants. In addition, it allowed us to investigate our second research question RQ2.

As explained in the Background and Related Work section, pilot interviews with decision makers revealed a preference for regular or grid layouts. As such both methods were therefore chosen in order to create those. Although the algorithms generate hexagonal grids, to maximize connectivity between concepts and space utilization, they can be modified to other tessellations as described by Emmer [18].

Both methods will use two common pieces of information for the calculation of the layouts and for development of the explanations: 1) the individual units or concepts, which represent the main components inside the maps; and 2) the relationships or similarities between the concepts, which help structure the layout and position the concepts relative to each other in the map.

A Reductive Explanation: IsoMatch

For the reductive approach, we decided to implement a projection method based on IsoMatch [20]. This technique is a suitable representative example of the current state of the art in creating layouts on a regular grid using dimensionality reduction.

This method consists of two phases. The first uses a dimensionality reduction algorithm to project the concepts into an arrangement in 2D space using the similarity information. For this phase we used Isomap [55] as it was the method suggested in the original Isomatch paper (Fig. 2a and 2b). The second phase uses an assignment algorithm to place the projected concepts to grid cells at a minimum movement cost. For this we generated a hexagonal grid in the same plane (Fig. 2c) and used the Kuhn-Munkres algorithm [38] for the assignment.

The generation of the grid was determined by the inter-quartile ranges of the projected concept points (defining which area has a greater concentration of points) and the number of cells needed (which can be greater than or equal to the number of concepts).

This projection algorithm produces consistent and aesthetically pleasing maps (Fig. 2d) representing the concept and similarity information. However, we believe projection algorithms are not easy to interpret or explain to non-technical participants, given the challenging concept that is highly-dimensional space.

Although we use Isomap as the basis for the projection, we designed the projection DDE to be representative of any dimensionality reduction technique, and therefore not represent steps involving the computation of dimensions. The projection method DDE would then observe these following steps, in which we will adhere to our previously defined requirements (needs to be visual, interactive and data-driven):

a. Display a scatter plot matrix. Each scatter plot represents the combination of two dimensions laying out the concepts on a plane, each row and column representing a dimension. (Fig. 2a).

b. Allow a first action which highlights the dimension that holds the least variance, i.e. the dimension with the least spread of concepts. Allow a second action that would remove the highlighted dimension from the scatter plot matrix. Allow participants to iterate this step until there are only two dimensions left (Fig. 2b)

c. Remove the matrix representation, and display the final projection of concepts along with the generated grid. Allow a third action to assign the concept points to grid cells one by one (Fig. 2c).

d. Display the final concept map layout, and remove the underlying grid (Fig. 2d).

A Constructive Explanation: Agglomerative Clustering

To contrast with the reductive explanation presented above, we decided to implement a constructive approach, i.e. an approach focusing on building relations between concepts, as described by Novak [40], which gradually introduces complexity into the map structure. In contrast to a projection method, which shows a complex multi-dimensional space reduced to two dimensions, here the concepts can be shown in a single dimensional space (i.e. a list of items) augmented to two dimensions, making it more intuitive to users. We also wanted to incorporate clustering as a basis, in order to reduce cognitive load for users. We therefore implemented this constructive method using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm [51], with a complete-link clustering [30], that will create a hierarchy of concept clusters.

In essence, the concepts are positioned directly relative to each other using this hierarchy. The most similar pair of items is picked first and positioned next to each other (Fig. 3b). For the rest of the process this pair of items is considered as a single entity, being translated and rotated together. This is repeated until we have the final map. When two groups of items are joined, they are positioned on the map grid in a way that minimizes the size of the map with respect to the original similarity measures in order to keep coherency (Fig. 3c). The final map will be complete once all groups are joined together (Fig. 3d).

The Agglomerative DDE follows these steps, again adhering to our previously defined three requirements (needs to be visual, interactive, and data-driven):

a. Display the concepts (represented by hexagons) vertically and in dendrogram order (from the linkage table). Note, the dendrogram is not shown at this stage (Fig 3a).

b. Allow a first action, highlighting the two most similar sets of concepts (only two concepts in the first stage). At the same time, display the dendrogram segment which joins those items. Allow a second action merging the two highlighted sets of concepts by positioning their hexagons together as they will appear in the final hex map. The merged items then move to the dendrogram node joining them (Fig. 3b).

c. Allow the iteration of step 2, joining and merging groups of concepts, until the final map is created (Fig. 3c).

d. Lastly, hide the dendrogram and zoom on the final map (Fig. 3d).

Data

The concepts used and displayed in this study were generated using topic modelling, specifically using Latent Dirichlet Allocation (LDA) [8]. In addition to the concepts themselves, we established the relationships between each of the concepts by calculating their similarities using the cosine distance of the topic distribution over documents [54]. The similarity matrix, together with the topic summaries, formed the main input for the two layout algorithms.

All of the topic models used for this study were generated using grant data from a national research directorate, downloaded via a publicly accessible online portal [1].

------

[1]Research Councils UK, Gateway to Research, **http://gtr.rcuk.ac.uk**, data retrieved in June 2016, last accessed 16th August 2017
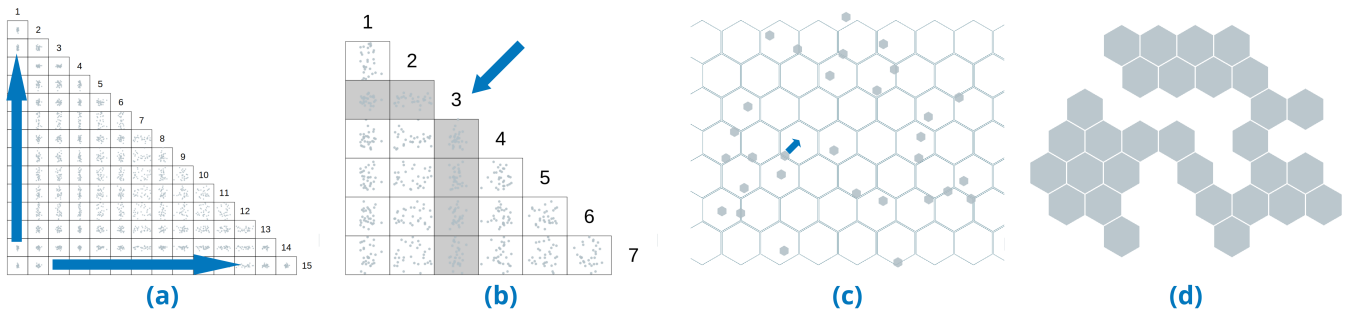
Figure 2. An illustration of the explanation for the projection layout method (reductive). (a) Displays a scatter plot matrix representing multiple dimensions. (b) Demonstrates the iterative action for participants removing dimensions with the least variance. (c) Shows the final projection and assignement of concept points to grid cells. (d) Presents an output of the projection method.
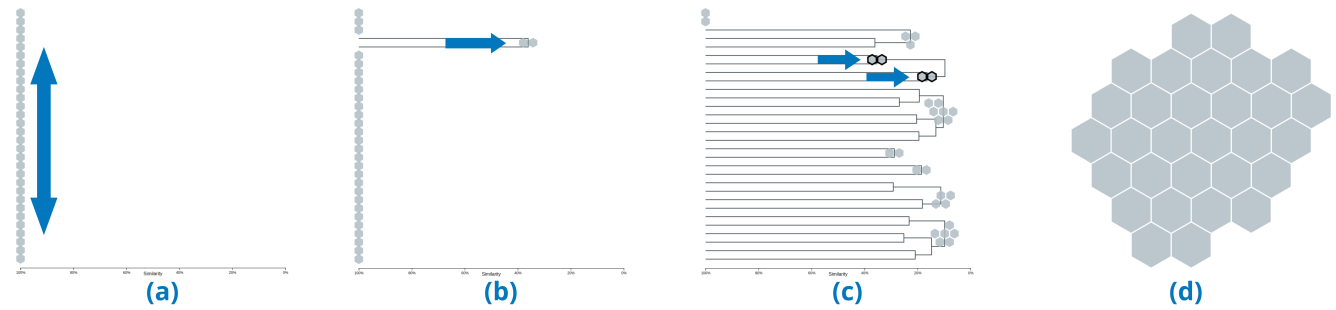
Figure 3. An illustration of the explanation for the agglomerative layout method (constructive). (a) Shows the ordering of the concepts by similarity using hierarchical clustering. (b) Demonstrates the actions of highlighting similar concepts and merging them together. (c) Displays the iterative process after 20 steps merging the various groups into a concise concept map. (d) Shows an output of the agglomerative method.

Pre-Study Work

As we are not investigating whether or not users can discriminate between materially accurate or inaccurate concept maps, we carried out a numerical analysis to compare how well each layout fits its underlying similarity information. For each of 125 different similarity matrices, inferred from topic models with varying sizes (10 to 50 items) and using different subset of the document pool, we created one concept map per layout method. We then computed the mean squared error of each concept map against the similarity matrix it represented.

An independent t-test revealed that, on average, the mean squared errors of the agglomerative concept maps (M = 0.213, SE = 0.002) were not significantly different from the mean squared errors of the projection concept maps (M = 0.209, SE = 0.002), t(248) = 1.171, p = 0.243, r = 0.074. This analysis showed, therefore, that there is no significant difference in the validity (or fitness), when measured numerically, of the two layout methods. These results would also be used during our study to reassure participants of the suitability of both methods.

Next we conducted a preliminary pilot study. It was essentially task-driven and aimed at providing both quantitative and qualitative results. The quantitative stages had several repeated tasks for the participant to focus on two highlighted concepts and understand their relative position within the DDE and then report their confidence (on a Likert scale) in being able to explain the position of the two highlighted concepts. Time taken on each task was also recorded. The final qualitative stage comprised of a questionnaire with both opinion and open-ended questions.

On evaluating the results of this pilot, we realised this study configuration did not allow us to access the deeper insights we were seeking. We therefore decided that a qualitative approach using in-depth semi-structured interviews would be more effective, as described by [2].

Study

Our qualitative study design would use a three-phased scenario-based semi-structured interview. During each phase of the interview participants would experience and interact with concept map and DDE applications. The questions and responses would be freely interleaved with interaction allowing deep interrogation of their views during the immersive scenario-based conversation. The semi-structured format would allow unexpected participant views to be pursued as appropriate and as they arose. In the rest of this section we describe the scenario that would be posed and then in subsections we describe each of the three interview phases.

To bring focus and context to the participants' opinions we would propose a scenario (or vignette [29, 4]) and ask them to place themselves within that scenario [60] in which their views would be sought. The concept maps embedded within the stimuli would represent research areas. The scenario would be described to participants as follows: "*The University commissioned us to create overviews of topics in order to facilitate a possible restructuring of research groups, by merging or splitting them. The data used for those maps was provided by a national research directorate. The restructuring could affect*

PhD or fellowship allocations and maybe the courses taught. Once the decisions are made your role is to announce them to affected groups". We planned to source our participants within our university. Therefore, featuring research and courses in the scenario was intentional to help them relate to it, and bring intrinsic and empathic motivations to the scenario [5].

After being told the scenario, participants would then be presented with a sequence of concept map applications, then they were invited to interact with these and discuss them in the context of the scenario. Each interview consisted of the three phases.

Interview Phase One: No DDE

During the first phase of the interview participants would be presented with a concept map (using either layout method, balanced across participants) and no DDE. The experimenter would point out a conflicting concept position (suggesting an unusual merging or splitting of research areas), and then let them explore the map for as long as they wished. Then the participant would be asked, given the scenario, about whether or not they agreed with the proposed research area splitting or merging, how they would communicate the decision to others and about their confidence in being able to explain the concept map. Follow-up questions would further probe their thoughts on the interface and the concept map, usability issues, using the concept map to make decisions, and about having the concept map layout process appear to be opaque (or a black-box).

### Interview Phase Two: DDE for Method X

In the second phase the participants would be presented with a concept map (again using either layout method, balanced across participants) including a DDE for its layout method. As in Phase One, the experimenter would point out a conflicting concept region, and then leave them to explore the concept map and the DDE for as long as they wished. Again, as in Phase One, the experimenter would ask questions based on the scenario. Additional questions would probe the participant's understanding of and feelings toward the layout method in the light of using the DDE. Any misunderstanding of the explanation would be corrected by the interviewer. Further questions would explore their confidence in explaining the layout to others contrasting this with Phase One, and usability would also be probed.

Interview Phase Three: DDE for Method Y

Finally, in the third phase, participants would be presented with a concept map (using a different layout method than Phase Two) and a DDE for its layout method. Similar to Phases One and Two, the experimenter would point to them a conflicting concept position, and then leave them to explore the concept map and the DDE. Then, the experimenter would probe their views with a selection of questions similar to Phase Two. Further questions would aim at prompting participants to contrast both layout methods and their confidence towards explaining them to others.

PROCEDURE

In this section we describe the procedure we followed to conduct our study. First we introduce the applications we de-
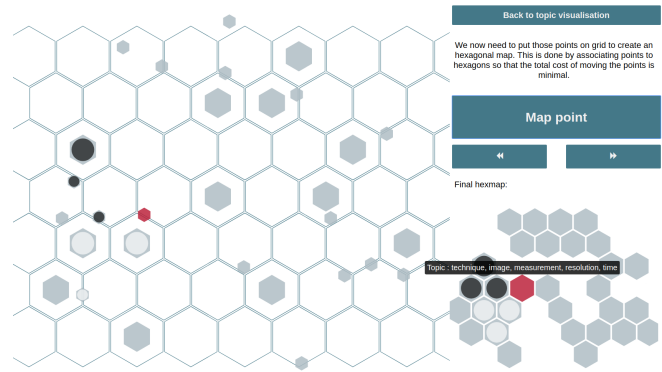
Figure 4. Screenshot of the application used in our study, here representing the projection data-driven explanation. The left hand side presents part of the main explanation visualization of a specific concept map. The top right hand corner shows a brief explanatory text regarding the next available action along with buttons allowing navigation through the explanation process. The bottom right hand corner displays a minimap of the final concept map to aid locating concepts in the explanation.

veloped, including data generation. Then we report on the interviews we conducted. Finally we describe the codebook used to categorize and analyze the interview data, after explaining our process in building it.

Applications

Two topic models were generated, both comprised of 30 topics and used the same number of documents (7,000); however both topic models used different subsets of the total document pool. Four maps were then produced, two from each topic model using the two layout methods. The maps produced from the first topic model were used for the first phase of each interview. The maps produced from the second topic model would, in addition, have the data for their data-driven explanation generated, and were used during the second and third phases of each interview.

We developed a web application to present concept maps with their DDE, using JavaScript and D3 to build interactive interfaces that would scale to any map data. In the concept map interface, the application would show concepts labelled by five words summarizing a topic. Each concept could be further interrogated by selecting and viewing it expanded with more details, such as a word cloud illustrating the topic and a list of the ten most similar concepts (Fig. 1). A button was included to access the concept map DDE interface.

The DDE interface consisted of three main elements: 1) the DDE visualization, which would change according to the layout method used to generate the concept map; 2) a mini-map of the final concept map; and 3) three buttons allowing the user to progress through the explanation.

The main button stated the next step available to the user: highlight the most similar items, or join the highlighted items on the dendrogram when displaying the agglomerative DDE; or highlight and then discard a dimension in the scatter matrix, then map points one by one onto the generated grid when displaying the projection DDE.The two other buttons would allow the user to automatically go forward or backward through the explanation.

By hovering over a concept, either on the explanation or the mini-map, the user would be able to highlight it in both visualizations, while simultaneously displaying a tooltip stating the top five labels of a topic. This would facilitate locating a concept in different parts of the view. The user would also be able to mark concepts on both visualizations by clicking them, allowing tracking of concepts throughout the explanation. Two groups of concepts could be tracked simultaneously (Fig. 4).

The application would be fully used in phases two and three of the interviews. For phase one of the interviews the button accessing the DDE would be hidden.

Interviews

We conducted the semi structured interviews with 10 participants, four females and six males, aged between 20 and 38 year old, recruited using convenience sampling and purposeful sampling based on gender to achieve as close as possible a gender balance given the participant pool [45, 61]. All participants were rewarded with a $12 Amazon voucher. Demographic information gathered on their background in data visualization established that while they spanned a range of backgrounds and skills in visualization none were experts, fitting our knowledgeable non-technical target users. Appropriate ethical approval and consents were obtained and all data was anonymized and unlinked. Prior to the interview, participants were also given a basic introduction to topic modelling, to give context to the data they would see.

The semi structured interviews were conducted following the scheme described in the study design. We allocated the layout methods presented to the participants in order to balance two aspects: 1) the layout method presented in phase one; and 2) the layout method presented in phase two (thus phase three) independently of phase one. The interviews took less than 45 minutes. Audio recordings were made and transcripts produced ready for coding.

Coding

We used computer-assisted qualitative data analysis software for coding, allowing us to effectively categorize information from the interviews, to count occurrences of certain types of comments and to calculate interoperator agreement statistics [50].

Transcripts were coded by two coders (author and a senior researcher), using a grounded theory (or inductive) approach [53], and both coders coded all transcripts to help ensure consistency and that nothing was missed [6, 11]. First the two coders read the same two randomly selected transcripts. Then a codebook which reflected the interview structure, common questions, and issues raised was developed. After coding the first interview transcript, the two coders met to discuss disagreements in coding and to refine codebook definitions [12]. The remaining transcripts were then coded accordingly. As coding continued an open coding approach was maintained and further codes were added [14]. A second pass through the data was made to ensure consistent coding of these further codes. This process is similar to that described by [17].
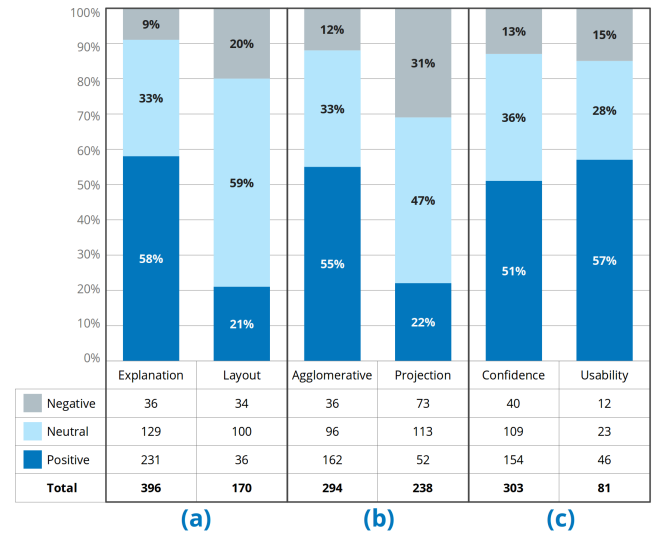
Figure 5. Coding frequencies of positive, neutral and negative sentiments for each of six categories: **Explanation** (discussion after participants had been provided with data-driven explanations); **Layout** (discussion prior to participants being provided with data-driven explanations); **Agglomerative** (concerning discussion specific to the agglomerative, constructive explanation); **Projection** (concerning discussion specific to the projection, reductive explanation); **Confidence** (concerning any discussion with regard to a participant's confidence); and **Usability** (concerning any discussion with regard to usability of the concept map and explanation interfaces).

At this stage, both coders found saturation, validating the sample size of the study. The coding comparison revealed a Cohen's Kappa figure of 0.54, with agreement between coders of 88.56% across all nodes and transcripts. These figures firstly, reflect a codebook which was meaningful (objective) but not too restrictive (allowed subjectivity) and secondly, reflect variations in the amount of context included with items during coding as is expected when coding semi-structured interviews with more than one coder [12].

The final codebook enabled the capture of 6 key categories of statements relative to: the data-driven **Explanation** of concept maps (regardless of method), the **Layout** of concept maps alone, the **Agglomerative** explanation, the **Projection** explanation, participants' **Confidence**, and the **Usability** of the map and interface. Each of these categories contained subcategories reflecting if the statements were either positive or negative. Other categories were also captured, such as personal reasoning, scenario reasoning, suggestions or questions, which contributed to a better understanding of our results.

RESULTS AND DISCUSSION

To quickly understand our results, we will first describe them briefly using a quantitative overview of our coded interviews. We will then address the results liked to our two research questions posed in the Aims section, RQ1 (overall effects of explanation on confidence) and RQ2 (specific effects of reductive and constructive approaches on confidence).

Overview of Interviews

As described above we coded interviews with participants into six key categories. For the purposes of discussion we divide these into three groups as shown in Figure 5 and referenced

below. The total references per category varied from 81 to 396 references. We acknowledge that the interview structure did influence the total discussion of each of the categories and therefore display sentiment proportions in the stacked bar charts (but in addition provide raw occurrence figures below) in Figure 5.

a. Examining Figure 5a shows that the proportions of positive sentiment expressed by participants rose from 21% with respect to discussion purely concerning concept maps, to 58% when data-driven explanations were discussed.

b. Figure 5b shows that participants were more positive (55%) when discussing the agglomerative (constructive) explanation than when discussing the projection (reductive) explanation (22%).

c. Figure 5c shows that while participant were explicitly invited to discuss usability (81 total references) in general they did not expand further on their in initial responses. In contrast confidence was discussed more frequently (303 total references) with only 13% negative comments.

Effect of Data-Driven Explanations on Confidence
In this part of the analysis we focus on RQ1: *What are the effects of data-driven explanations on confidence?* First we examine the views expressed in the first phase of the interviews when participants only had a concept map with no explanation. Then we describe three subthemes arising in the second phase of the interviews when participants had a data-driven explanation accompanying the concept map.

Confidence and Concept Maps
In the first phase of the interviews, when asked about their level of confidence towards concept maps, four of the ten participants expressed positive confidence towards the layout of the concept map: "*I think by showing them the visualization, and show the more relation they have to this field, I would be confident to support my opinion.*" (P4); "*this interface will give me more confidence, for areas that I don't know much about.*" (P1).

However, some reservations were also expressed. (P5), for example, said they needed more data-based evidence: "*... directly like that it would not be too clear to me how this grid was constructed. ... There needs to be a stronger relation to the underlying data*". Others, like (P7), clearly expressed a need for explanation: "*I think it would be useful to have some sort of secondary explanation on top of that [concept map] because I think a lot of people ..., would find it difficult to find the correlation*". A reason for this need was expressed by (P3): "*I think I would need more time, to work out exactly what is going on. ... before I actually had to explain it to someone*".

When asked to explain a concept map or the decision it led to within the scenario, participants were uncertain, thinking it required more time to figure out why a situation was represented that way, and asking for more data-based evidence and explanation. We believe this result reinforces the need for data-driven explanations and their influence in confidence.

Confidence and Data-Driven Explanation
This theme was exposed from the analysis of phase 2 of the interviews when participants experienced a data-driven explanation. We have subdivided this theme into three categories to aid the description of the various influences on confidence which we have uncovered: **evidence**, **interactivity**, and **adoptability**.

Evidence: When exposed to data-driven explanations it was noted that a majority of participants (eight out of ten) plainly indicated having stronger confidence in being able to explain a given concept map to others: e.g. "*With knowing how this thing is constructed, I am more confident that the splitting and merging ... makes sense to me.*" (P5) and "*Now you know what you are going to say and why you are going to say it. It does help to know*" (P6).

Participants commented that the DDE made the concept map appear more robust: e.g. "*The help [explanation] is like a bonus, to trust the map more ... it's better than just the map by itself*" (P8); "*It is more realistic, actually*" (P6); "*[the explanation] gives me a second level of information, which the topic map by itself lacks, so I can see all those similarity values, or strength so to say, which I cannot see from the topic map itself*" (P5).

(P5) points out that the data-driven explanations provide evidence to be used in any self-explanation that they might be called on to give to others in their decision making role in the scenario. Concept maps alone had provided less evidence. This opinion was shared by eight of our participants, and three directly related it to an increase of confidence in their ability to explain decisions: e.g. "*... The explanation gives a good view of how the system works and how, for example, the merging or the breaking [of areas] is succeeded*" (P4); "*... it makes it a lot clearer, how the decisions are founded because A and B are connected and C and D are kind of neighboring. ... It would give you more confidence in decisions that are made ... if it was backed with like hard evidence there*" (P7); "*... the previous one [concept map alone] it was just blind and I made my decision based on my feelings rather than knowing what I am doing, but here I did have some material to base my decision on.*" (P6).

> **R1**: We recommend that designers incorporate data-driven explanations into their visualizations, as we have found that explanations particular to the underlying data: increase participants' perception of the robustness of the layout process, increase participants' understanding of how and why the layout was organised in the way presented, and also improve participants' confidence in their ability to explain the layout process to others.

Interactivity: Having the explanations built to be interactive helped in improving participants' confidence. For example, (P5) declared "*It is definitely good to have the explanation where you can pause, play back and play forward, so that I can actually pause at points that I might be interested in. I mean if I really would have to decide something, then I personally would be really interested to see that visually and to find things*

where I have to argue, and see them kind of over time.". This extended to the data-driven aspect as well. Eight participants said it made the explanation more meaningful: e.g. "*I think it's better to have it showing why it came to this specific example rather than a generic one*" (P10); "*It is better to have a layout [explanation layout] that is related to the data in real time. ... I would be more confident to have a layout [explanation layout] that is directly related to the data*" (P8).

> **R2**: We found that incorporating interactivity in explanation increased user confidence and engagement as it enabled participants to interrogate the process and understand the information at their own pace. We believe this increase is particularly pronounced when compared with the use of non-interactive media, such as video tutorials. We therefore recommend that interaction is incorporated in explanation mechanisms in order to allow users to query individual items and have control over the step-by-step process.

Adoptability: Seven out of ten participants explicitly expressed their willingness to reuse data-driven explanations, which attests to a stronger confidence in concept maps. For them this relates to it providing evidence and being interactive: e.g. "*If ... you can show the explanation, then I think it is easier to convince people that this is correct. Without the explanation ... it might be not that clear to people how this map is generated.*" (P5); "*I have a strong explanation of how the results were produced and by explaining to everyone, ... I think that no one could ... argue*" (P4); "*I think it would resonate more with people, having this kind of visualization [explanation], kind of hard evidence ... when someone is talking to you about something, I don't think it does ring as true as kind of having it shown right in front of you*" (P7).

We found that participants expressed willingness to adopt and make repeated use of the explanations as the associated interactions provided them with the confidence to explain, argue, and present concept maps in detail. This result provides further motivation for the adoption of interactive data-driven explanations as discussed in R1 and R2.

Effect of Explanation Mechanisms on Confidence

In this part of the analysis we address RQ2: *What are the specific effects of reductive and constructive approaches on confidence?* Although we were interested in issues concerning algorithmic interpretability, credibility also emerged to be strongly related to this research question. Both explanation mechanisms were found to be credible, interestingly however two views arose which discriminate the mechanisms in more detail. First, there is the credibility of the concept map itself, in which **density** played an important role. Secondly, there is the credibility of the explanation, where **clarity** was the main factor.

Effect of Map Density

In the case of the projection method, the packing of the final layout seemed to make the process more credible improving confidence in the participants, compared to the agglomerative method, as expressed by (P5): "*I would say ... second result [projection], seems to be more in line with what I would expect,*

because you actually don't have a dense map, so you have missing grid line in between which would translate to me into having bigger distances. So the second map actually includes the notion of distance between topics better than the first one".

This density of the map is directly due to the inner workings of the layout algorithms. The projection method values the overall difference between items, spreading out the items on the final map. The agglomerative method gives more importance to close similarities creating a dense map. Here (P8), talking about the agglomerative method, exposes this issue: "*At the beginning ... all the topics that are together are really [together], we can understand that they are related together. But at one time ... we force them together even if they don't have any links. ... I feel it's unnatural*".

> **R3**: We recommend designers use layout techniques that enable them to control the density of the maps as our results showed that while dense layouts provide efficient screen utilization, sparser layouts make the relationships between concepts more obvious and consequently increase users' confidence.

Effect of Process Clarity

In the case of the agglomerative method, the clarity of the process made it more credible, compared to the dimensionality reduction process which appeared confusing. "*Just looking at the explanation that you provided I would say that the clustering [agglomerative] gives me a clearer view of how the topics are related to each other, and how the process is done*" (P4); "*To be honest I am a bit confused by this one [projection]*" (P9); "*I understand the way it [agglomerative] works more, ... merging them together one by one, by the most similarity first ... . I think that makes better sense ... than the previous process [projection]... .*" (P10).

This was explained by participants by the fact that the agglomerative method is more in line with their expectations and that the projection method requires prior knowledge: e.g. "*It resounded more to me. ... It was just the clearest ... It visually it made more sense*" (P7); "*I found the previous one [agglomerative] shows the clustering more naturally as how I think. And this dimension stuff [projection] is, I have too many questions and too many things I need to understand first before I can fully understand the ... why this is presented this way.*" (P9).

The agglomerative method was also found by nine participants to be more interpretable. This was reflected by the number of participants to which the projection method had to be clarified by the interviewer (seven) compared to the agglomerative method (one). Most of the time when talking about understanding, participants stepped out of the scenario and spoke from a personal perspective: e.g. "*[It's] Easier to understand with the dendrogram*" (P3); '*I did understand this one [agglomerative] better than the previous one [projection]*" (P8).

However, often participants also related to the scenario, and expressed their opinions regarding a larger audience: e.g. "*I think in terms of explanation for the audience it [agglomerative] would be more understandable to people*" (P4); and

"The previous one [projection] is a bit complex. ... This one [agglomerative] should be accessible to everybody." (P6). This was also agreed by (P2), who personally had a preference for the projection method: *"It [agglomerative] requires less explanation perhaps. ... This is easier for a more generic audience to follow."*.

This preference for the agglomerative method continued when we asked participants about their confidence in their ability to explain the process or a decision made from the map to other people. All participants agreed the agglomerative method would be easier to explain: e.g. *"I think if you feel more confident in the system, you are going to be more confident with the decision or communicating the decision. ... I would find that [agglomerative] more useful"* (P7); *"Definitely this one [agglomerative] would be easier to explain"* (P8); and *"I feel like I would not be able to explain it [projection] myself"* (P1).

Six of the ten participants explicitly said they would reuse the agglomerative method in preference to the projection method and eight stated they had more confidence when using the agglomerative method than with the projection method.

Participants were sensitive to the differences in the processes and this can be illustrated with the thoughts of (P10): *"In the other one [projection] ... you found them [dimensions] and they disappeared, and then nothing really ... Whereas in this one [agglomerative] ... they [the concepts] merged together and kind of took their own shape ... You can understand a bit more how those topics were in relation to each other"*.

The agglomerative explanation was designed to gradually increase in complexity, generating a structured map one step at a time, from a simple list of concepts. Looking at our evidence in detail, we believe that adding one piece of information at a time helped build process credibility and interpretability. Conversely, the projection method starts as a complex state containing the majority of the information, that is then gradually simplified. Although the process of elimination is natural for some people, we observed it to be confusing for the majority of our participants. In particular, we believe that presenting multiple possible structures, before discarding all but one, made the process less engaging for our participants.

> **R4**: We recommend that designers adopt constructive as opposed to reductive layout methods, as participants found the agglomerative approach was both clearer to understand and considered it a more natural process. We believe that this is due to the gradual presentation and aggregation of concepts which means that participants are not required to maintain large numbers of ideas in working memory at one time. This reduces confusion in users and gives them more confidence in their ability to understand and explain the resulting concept maps.

CONCLUSION AND FUTURE WORK

In this paper we report the effect of data-driven explanations on users' confidence both in terms of their ability to understand, and explain, the layouts of automatically generated concept maps. To this aim we conducted a qualitative study using in-depth scenario-based interviews that exploited interactive, visual, explanations of *constructive* and *reductive* layout methods.

During these interviews, participants used and discussed concept maps that were provided with and without our data-driven explanations. Our participants reported having stronger confidence when they used the data-driven explanations, as they provide case-specific evidence and interactivity that allows both control over the explanation's pacing and the ability to query the underlying evidence (R1 and R2). These results were further reinforced by a frequency analysis which showed that users were proportionately more positive when discussing data-driven explanation, than when commenting on concept map layouts on their own (58% vs. 21% respectively).

The participants were also provided with two different types of explanation, one for each of the layout methods. The *reductive* approach was based upon standard projection methods, while the *constructive* approach was derived from a simple agglomerative clustering technique. Our aim in providing these two methods was to enhance the diversity of the stimuli rather than perform A/B testing, as the latter would not be possible without confounding the effect of the two layout and explanation methods.

The two different layout methods discussed above revealed two important design considerations that can affect the credibility of concept maps (R3 and R4).

First, that layout density affects users' perception of the *clarity* of the map, as it alters their ability to perceive structure (as dense packing prevents easy abstraction by users of concepts into groups). We therefore recommend that designers choose algorithms which provide variable packing densities so that they can trade screen real estate for the communication of structure and perceived clarity (R3). It should be noted that while our implementation of the agglomerative method produced dense maps in this study, this is not intrinsic to the method. In future work we plan to investigate different ways of communicating inherent structure within the map's similarity data.

Second, the two types of stimuli exposed participants' strong preference for *agglomerative* or *constructive* approaches as opposed to the *reductive* projective methods in which a complex problem is presented to users and then information is gradually discarded. In particular, the repeated aggregation of pairs of concepts into clusters in the first approach means that the user's working memory is not overloaded, which we believe greatly contributes to users' perception of simplicity and clarity (R4).

To conclude, we believe this study also highlights the need to further understand and research the underlying issues that affect user confidence when generating various visualizations using automated tools for planning, decision-making and collaborative activities.

REFERENCES

1. Gregory A Aarons, Rebecca S Wells, Karen Zagursky, Danielle L Fettes, and Lawrence A Palinkas. 2009. Implementing evidence-based practice in community mental health agencies: A multiple stakeholder analysis. *American journal of public health* 99, 11 (2009), 2087–2095.

2. Anne Adams and Anna L Cox. 2008. Questionnaires, in-depth interviews and focus groups. In *Research Methods for Human-Computer Interaction*, Paul Cairns and Anna L Cox (Eds.). Cambridge University Press, Chapter 2, 17–34.

3. Lynda A. Anderson, Margaret K. Gwaltney, Demia L. Sundra, Ross C. Brownson, Mary Kane, Alan W. Cross, Richard Mack, Randy Schwartz, Tom Sims, and Carol R. White. 2006. Using Concept Mapping to Develop a Logic Model for the Prevention Research Centers Program. *Preventing Chronic Disease* 3, 1 (2006).

4. Christine Barter and Emma Renold. 1999. The use of vignettes in qualitative research. *Social research update* 25, 9 (1999), 1–6.

5. Victoria Bellotti, Alexander Ambard, Daniel Turner, Christina Gossmann, Kamila Demkova, and John M Carroll. 2015. A muddle of models of motivation for using peer-to-peer economy systems. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1085–1094.

6. John Benson and Nicky Britten. 2002. Patients' decisions about whether or not to take antihypertensive drugs: qualitative study. *Bmj* 325, 7369 (2002), 873.

7. David M Blei. 2012. Probabilistic topic models. *Commun. ACM* 55, 4 (2012), 77–84.

8. David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

9. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2301–2309.

10. Kevin W Boyack, Richard Klavans, and Katy Börner. 2005. Mapping the backbone of science. *Scientometrics* 64, 3 (2005), 351–374.

11. Laila Burla, Birte Knierim, Jurgen Barth, Katharina Liewald, Margreet Duetz, and Thomas Abel. 2008. From text to codings: intercoder reliability assessment in qualitative content analysis. *Nursing research* 57, 2 (2008), 113–117.

12. John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research* 42, 3 (2013), 294–320.

13. Emilio Carrizosa, Vanesa Guerrero, and Dolores Romero Morales. 2017. Visualizing proportions and dissimilarities by space-filling maps: a large neighborhood search approach. *Computers & Operations Research* 78 (2017), 369–380.

14. Juliet Corbin and Anselm Strauss. 2008. Basics of qualitative research: techniques and procedures for developing grounded theory. 2008. (2008).

15. Dan Cosley, Shyong K Lam, Istvan Albert, Joseph A Konstan, and John Riedl. 2003. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 585–592.

16. Martin Davies. 2011. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher education* 62, 3 (2011), 279–301.

17. Niklas Elmqvist and Ji Soo Yi. 2015. Patterns for visualization evaluation. *Information Visualization* 14, 3 (2015), 250–269.

18. Michele Emmer. 2005. The visual mind II. (2005).

19. Holly J Falk-Krzesinski, Noshir Contractor, Stephen M Fiore, Kara L Hall, Cathleen Kane, Joann Keyton, Julie Thompson Klein, Bonnie Spring, Daniel Stokols, and William Trochim. 2011. Mapping a research agenda for the science of team science. *Research Evaluation* 20, 2 (2011), 145–158.

20. Ohad Fried, Stephen DiVerdi, Maciej Halber, Elena Sizikova, and Adam Finkelstein. 2015. IsoMatch: Creating informative grid layouts. In *Computer graphics forum*, Vol. 34. Wiley Online Library, 155–166.

21. Emden R Gansner, Yifan Hu, and Stephen Kobourov. 2010. GMap: Visualizing graphs and clusters as maps. In *Visualization Symposium (PacificVis), 2010 IEEE Pacific*. IEEE, 201–208.

22. Elena Garces, Aseem Agarwala, Aaron Hertzmann, and Diego Gutierrez. 2017. Style-based exploration of illustration datasets. *Multimedia Tools and Applications* 76, 11 (2017), 13067–13086.

23. Alyssa Glass, Deborah L McGuinness, and Michael Wolverton. 2008. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces*. ACM, 227–236.

24. Nitesh Goyal, Gilly Leshed, Dan Cosley, and Susan R Fussell. 2014. Effects of implicit sharing in collaborative analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 129–138.

25. Raja Gumienny, Steven P Dow, and Christoph Meinel. 2014. Supporting the synthesis of information in design teams. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 463–472.

26. Steven Hansen, N Hari Narayanan, and Mary Hegarty. 2002. Designing educationally effective algorithm visualizations. *Journal of Visual Languages & Computing* 13, 3 (2002), 291–317.

27. Gunnar Harboe, Jonas Minke, Ioana Ilea, and Elaine M Huang. 2012. Computer support for collaborative data analysis: augmenting paper affinity diagrams. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1179–1182.

28. Jason L Harman, John O'Donovan, Tarek Abdelzaher, and Cleotilde Gonzalez. 2014. Dynamics of human trust in recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 305–308.

29. Nicholas Jenkins, Michael Bloor, Jan Fischer, Lee Berney, and Joanne Neale. 2010. Putting it in context: the use of vignettes in qualitative interviewing. *Qualitative Research* 10, 2 (2010), 175–198.

30. Benjamin King. 1967. Step-wise clustering procedures. *J. Amer. Statist. Assoc.* 62, 317 (1967), 86–101.

31. René F Kizilcec. 2016. How much information?: Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2390–2395.

32. Bart P Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. 2012. Inspectability and control in social recommenders. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 43–50.

33. Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2119–2128.

34. Zachary C Lipton. 2016. The mythos of model interpretability. *arXiv preprint arXiv:1606.03490* (2016).

35. Paulo JG Lisboa. 2013. Interpretability in Machine Learning–Principles and Practice. In *International Workshop on Fuzzy Logic and Applications*. Springer, Cham, 15–21.

36. Duen-Ren Liu, Chin-Hui Lai, and Hsuan Chiu. 2011. Sequence-based trust in collaborative filtering for document recommendation. *International Journal of Human-Computer Studies* 69, 9 (2011), 587–601.

37. Wendy L Martinez, Angel R Martinez, Angel Martinez, and Jeffrey Solka. 2010. *Exploratory data analysis with MATLAB*. CRC Press.

38. James Munkres. 1957. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics* 5, 1 (1957), 32–38.

39. Joseph Novak. 1991. Clarify with concept maps. *The science teacher* 58, 7 (1991), 44.

40. Joseph D Novak and Alberto J Cañas. 2008. The theory underlying concept maps and how to construct and use them. (2008).

41. Joseph D Novak and Dismas Musonda. 1991. A twelve-year longitudinal study of science concept learning. *American Educational Research Journal* 28, 1 (1991), 117–153.

42. Conor Nugent and Pádraig Cunningham. 2005. A case-based explanation system for black-box systems. *Artificial Intelligence Review* 24, 2 (2005), 163–178.

43. Stefano Padilla, Thomas S Methven, David W Corne, and Mike J Chantler. 2014. Hot topics in CHI: trend maps for visualising research. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 815–824.

44. Stefano Padilla, Thomas S Methven, David A Robb, and Mike J Chantler. 2017. Understanding Concept Maps: A Closer Look at How People Organise Ideas. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 815–827.

45. Michael Quinn Patton. 1990. *Qualitative evaluation and research methods*. SAGE Publications, inc.

46. Wolter Pieters. 2006. Acceptance of voting technology: between confidence and trust. In *iTrust*. Springer, 283–297.

47. Wolter Pieters. 2011. Explanation and trust: what to tell the user in security and AI? *Ethics and information technology* 13, 1 (2011), 53–64.

48. Roberto Pinho, Maria Cristina F de Oliveira, and Alneu de A Lopes. 2009. Incremental board: a grid-based space for visualizing dynamic data sets. In *Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 1757–1764.

49. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1135–1144.

50. David Silverman. 2013. *Doing qualitative research: A practical handbook*. SAGE Publications Limited.

51. Robert R Sokal. 1958. A statistical method for evaluating systematic relationship. *University of Kansas science bulletin* 28 (1958), 1409–1438.

52. Justin Solomon, Gabriel Peyré, Vladimir G Kim, and Suvrit Sra. 2016. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 72.

53. Anselm L Strauss. 1987. *Qualitative analysis for social scientists*. Cambridge University Press.

54. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2005. Introduction to data mining. 1st. (2005).

55. Joshua B Tenenbaum, Vin De Silva, and John C Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. *science* 290, 5500 (2000), 2319–2323.

56. William MK Trochim. 1989. An introduction to concept mapping for planning and evaluation. *Evaluation and program planning* 12, 1 (1989), 1–16.

57. William M Trochim. 2017. Hindsight is 20/20: Reflections on the evolution of concept mapping. *Evaluation and program planning* 60 (2017), 176–185.

58. Tamara Van Gog, Liesbeth Kester, Fleurie Nievelstein, Bas Giesbers, and Fred Paas. 2009. Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior* 25, 2 (2009), 325–331.

59. Colin Ware. 2012. *Information visualization: perception for design*. Elsevier.

60. Kelly D Wason, Michael J Polonsky, and Michael R Hyman. 2002. Designing vignette studies in marketing. *Australasian Marketing Journal (AMJ)* 10, 3 (2002), 41–58.

61. Tom Wengraf. 2001. *Qualitative research interviewing: Biographic narrative and semi-structured methods*. Sage.

62. James A Wise, James J Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. 1995. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Information Visualization, 1995. Proceedings.* IEEE, 51–58.

63. Iwan GJH Wopereis, Paul A Kirschner, Fred Paas, Slavi Stoyanov, and Maaike Hendriks. 2005. Failure and success factors of educational ICT projects: a group concept mapping approach. *British Journal of Educational Technology* 36, 4 (2005), 681–684.