DOI: 10.1111/cgf.12880
Eurographics Conference on Visualization (EuroVis) 2016
K.-L. Ma, G. Santucci, and J. van Wijk
(Guest Editors)

Volume 35 (2016), Number 3

Comparing Node-Link and Node-Link-Group Visualizations From An Enjoyment Perspective

Bahador Saket[1], Carlos Scheidegger[2] and Stephen Kobourov[2]

[1]Georgia Institute of Technology, Georgia, USA
[2]University of Arizona, Arizona, USA

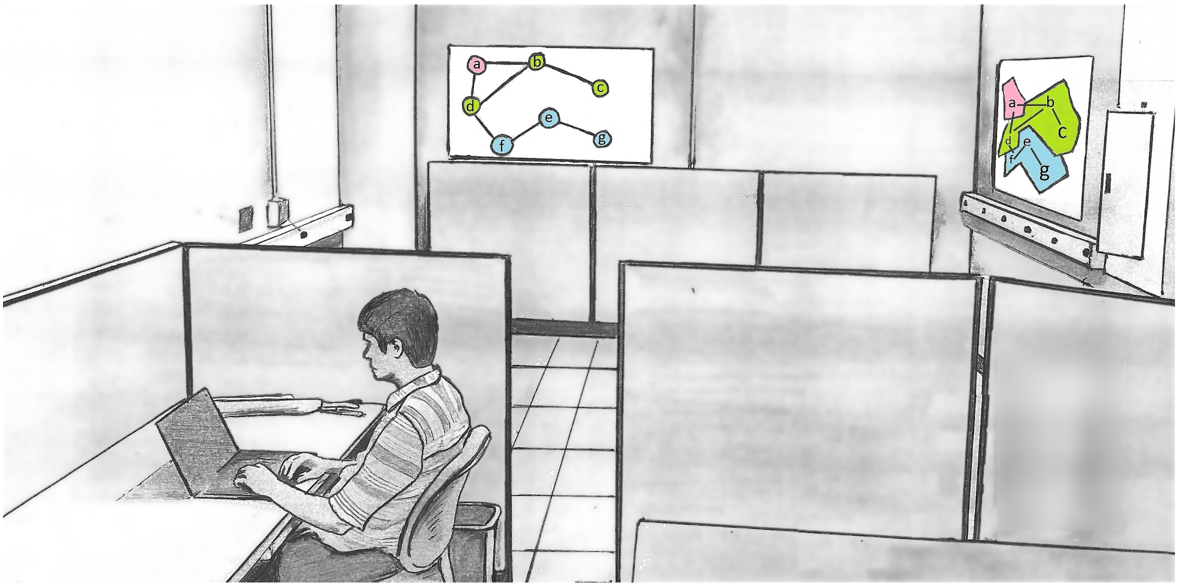

Figure 1: *An illustration of the experimental setting for comparing node-link and node-link-group visualizations from the perspective of participant enjoyment. In the first phase we studied the spontaneous interaction of participants with the posters installed in the back of the room. In this phase, experimenters are outside the room and participants are not aware this is part of the experiment. The second and third phases were conducted in the forefront area, where the laptop is located.*

Abstract

While evaluation studies in visualization often involve traditional performance measurements, there has been a concerted effort to move beyond time and accuracy. Of these alternative aspects, memorability and recall of visualizations have been recently considered, but other aspects such as enjoyment and engagement are not as well explored. We study the enjoyment of two different visualization methods through a user study. In particular, we describe the results of a three-phase experiment comparing the enjoyment of two different visualizations of the same relational data: node-link and node-link-group visualizations. The results indicate that the participants in this study found node-link-group visualizations more enjoyable than node-link visualizations.

1. Introduction

"What makes a visualization good?" is a question as old as the field itself. Within visualization, quantitative evaluation usually is focused on performance time and accuracy [SSK15]. More recently, there has been a concerted effort to take into account aspects beyond time and error. For example, the BELIV workshop is a well-known venue created to encourage the study of novel evaluation methods. For example, several recent papers study the memorability of visualizations [BARM*12, BVB*13, HKF15]. Other aspects, such as enjoyment and engagement, are not as well explored, even though enjoyment is often given as a reason to consume visualizations [BM13]. Moreover, we seek as a field to eventually understand the impact of enjoyment on performance; for example, positive mental states appear linked to better problem-solving performance in general [Fre98]

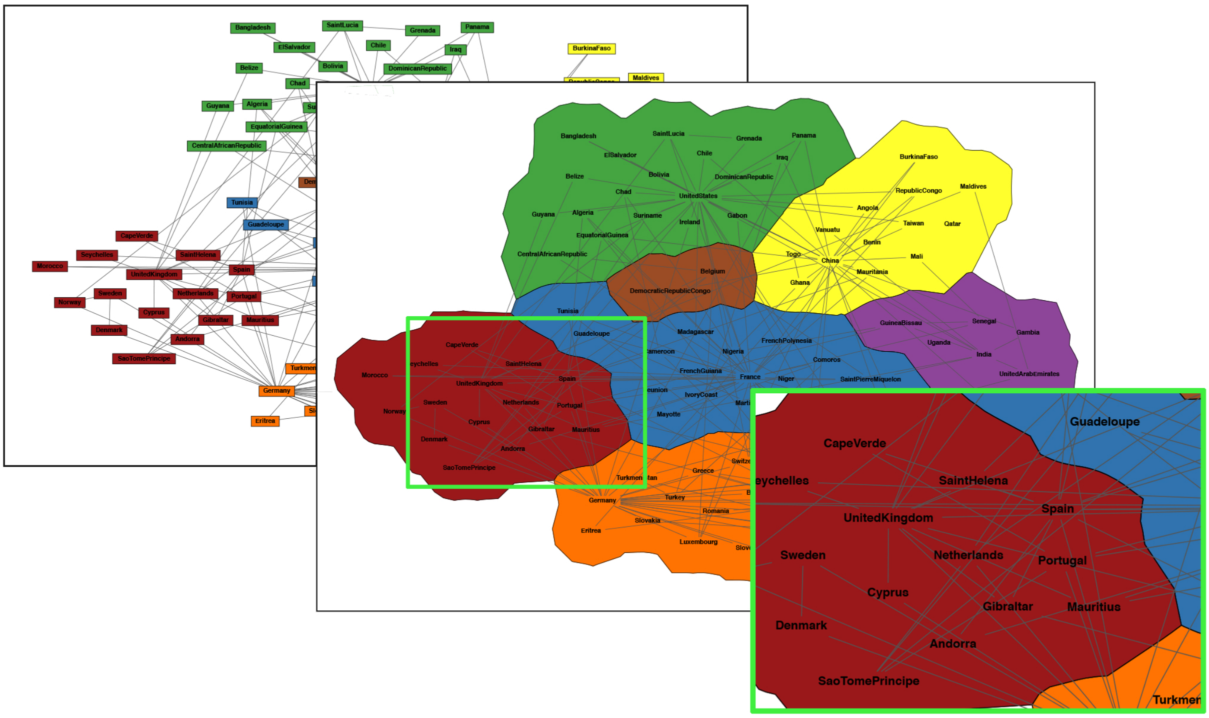

Figure 2: *Examples of visualizations considered in this study. We investigate the enjoyability of relational data represented with node-link (left-side) and node-link-group (right-side) visualizations.*

and in information visualization in particular [HSF*13]. If this effect turns out to be significant, then the enjoyability of experiences should be accounted for in the visualization design process.

In this paper we describe a multiphase experiment comparing two different visualizations of the same relational data: node-link (NL) diagrams and node-link-group (NLG) visualizations. The goal of the experiment is to study the extent to which different visualizations can affect the overall enjoyment of the participants.

NL diagrams are a standard visualization of relational data with the data objects represented by a dot or a circle in the plane and data relations between objects represented by a curve connecting the corresponding objects; see Fig. 2-left. Cluster information is encoded by coloring nodes in the same cluster with the same color. NLG diagrams are similar to NL diagrams but show cluster information by placing groups of nodes in colored regions of the plane, thus creating the impression of a geographic map; see Fig. 2-right. Many recent tools generate visualizations which can be described as node-link-group: BubbleSets [CPC09], LineSets [ARRC11], GMap [GHK*10b], and KelpFusion [MRS*13].

As part of an experiment measuring the memorability of embellished and plain charts, Bateman et al. [BMG*10] informally asked participants to rate the enjoyability of each type of chart using a Likert scale; most of these results were replicated in a recent study by Li et al. [LM14]. Both studies suggest that embellished charts are more enjoyable than plain ones, but "enjoyment" can encompass many different elements [Csi90]. Thus, although previous studies [BMG*10, LM14] find which visualization type is more enjoyable than another, by conducting studies that control for more specific variables we hope to eventually understand what makes one type of visualization more enjoyable than another.

We measure and compare the enjoyment of node-link and node-link-group visualization techniques with a three-phase experiment which we describe in detail in Section 3. The main goal in the experiment is to balance qualitative assessment with quantitative methods in order to measure enjoyment and engagement. Our main insight for measurement is that *if users are surreptitiously given the option to experience either of two visualizations, they will spend longer with the one they enjoy better*. We combine this with self-reported measurements of enjoyment using a recently proposed model [SSK15] based on matching Csikszentmihalyi's *flow* [Csi90] to Munzner's nested model for validation [Mun09]. Finally, we ask users open-ended questions about their preferences, as advocated for qualitative experiments [OK14]. As discussed in Sections 3 and 4, *the participants in this study found node-link-group visualizations more enjoyable than node-link visualizations,* and the difference is statistically significant.

2. Related Work

Experiments with NL and NLG Visualizations: When visualizing a relational dataset with a node-link diagram, people tend to place groups of nodes in distinct spatial regions that do not overlap with regions occupied by other groups, and edges in the group visually delineate the group itself, as shown by van Ham and Rogowitz [HR08]. Jianu et al. [JRHT14] assessed the effectiveness of four techniques for visualizing relational datasets in terms of response time and error. The main goal of the study was to match specific tasks to specific visualizations. They asked participants to perform 10 different tasks and their results indicate that BubbleSets outperform the other visualizations in tasks that involve group perception and understanding and that GMap might be more memorable than other visualizations. Also in terms of task accuracy and time, Saket et al. [SSKB14] investigated the effectiveness of three different relational data visualizations: point clouds (N diagrams), node-link diagrams (NL diagrams), and node-link-groups (NLG diagrams). Their results show that progressively adding more information (from just nodes, to nodes and links, to nodes and links and groups) does not necessary result in slower and less accurate task performance. For example, participants who used NLG diagrams performed network-based tasks as fast as with NL diagrams. More recently, Saket et al. [SSKB15] conducted an experiment to study the long-term memorability of the underlying data represented in NL and NLG diagrams. They found that subjects recall data in NLG diagrams more accurately. While these studies measured aspects such as performance time, accuracy and memorability, they did not address the relative enjoyment of NL and NLG diagrams, the problem we study in this paper.

Measurement of Enjoyment in Other Fields: Enjoyment has been carefully studied in psychology. One of the most well known models for understanding and measuring enjoyment in psychology is the flow model of Csikszentmihalyi [Csi90]. In a series of experiments in different countries people were asked to describe when and how they achieved the highest level of enjoyment when performing some activity. As Csikszentmihalyi writes, “Regardless of culture, social class, gender or age, the respondents described enjoyment in very much the same way. What they did to experience enjoyment varied dramatically - the elderly Koreans liked to meditate, the teenage Japanese liked to swarm around in motorcycle gangs — but they described how it felt when they enjoyed themselves in almost identical terms” [Csi90]. He then suggests several factors that encompass the experience of enjoyment:

- **Challenge:** the activity should be challenging and require skill
- **Focus:** it should be possible to concentrate on the task
- **Clarity:** it should be possible to concentrate on the activity *because it has clear goals*
- **Feedback:** it should be possible to focus on the activity *because it provides immediate feedback*
- **Control:** participants should feel a sense of control over actions
- **Immersion:** participants should lose the concern for self (this is sometimes described as being “in the zone”)

Rathunde and Csikszentmihalyi [RC05] used the flow model to understand experiences in two different educational settings: Montessori and traditional, indicating that Montessori students experienced enjoyment more frequently. The flow model was also used to study enjoyment in an interactive music environment [PA04]. In an experiment performed with professional pianists, a significant relationship was found between the enjoyment of the pianist and the pianist’s heart rate and blood pressure [dMTHU10]. Jackson interviewed twenty-eight elite-level athletes to understand how the flow (enjoyment) state is experienced by athletes [Jac96]. In order to develop a problem solving environment which support creativity, Vass et al. [VCS02] applied several theories, including the flow model. The flow model has also been applied to develop a framework for constructing engaging commercial websites [Jen00] and to assess information systems [Art96]. Since it is known that flow experiences are similar across different fields [Csi90], we can expect to observe and measure flow (and, consequently, enjoyment) in visualization in the same way.

Measurement of Enjoyment in Visualization: Elmqvist et al. [EVMJ*11] define fluid interaction in the context of information visualization as a concept characterized by smooth, seamless and powerful interaction; responsive, interactive, and rapidly updated graphics; and careful, conscientious, and comprehensive user experiences. A fluid information visualization interface has three properties: it promotes flow, supports direct manipulation, and minimizes the gulfs of action. Bateman et al. [BMG*10] conducted a study to evaluate the comprehension and recall of charts using an embellished version and a plain version. As a part of this study they also asked participants to rate the enjoyability of each type of chart. Their results suggest that embellished charts are more enjoyable than plain ones. Li et al. recently reported a replication, limiting their selection to those charts that consisted of data sets with 10 or more observations. Li et al. [LM14]’s study had a similar result: embellished charts were found to be more enjoyable than plain ones. In a very recent study, Haroz et al. [HKF15] assessed user engagement with ISOTYPES by measuring the total amount of time participants spent looking at different visualizations, similarly to the test we run in one of the phases of our study. Boy et al. [BDF15] investigated the effects of initial narrative visualization techniques and storytelling on user engagement by examining interaction logs (e.g., amount of time spent on exploration, number of meaningful interactions). Recently, Mahyar et al. [MKK15], Tanahashi et al. [TM15], and Saket et al. [SSK15] proposed models of enjoyment in visualization. In particular, Saket et al. considered different elements of flow (challenge, focus, clarity, feedback, control, immersion) and argued that these elements correspond to specific levels of Munzner’s nested model [Mun09].

Alternative methods for measuring enjoyment and engagement in visualizations have also been considered. Cernea and Kerren [CKE11] employed a mobile electroencephalographic headset for detecting emotional responses, when working with a visualization. Peck et al. [PYO*13] argue that functional, near-infrared spectroscopy is a viable technology for understanding the affect of visual design on a person’s cognition processes. Fabrikant et al. [FCPM12] measured the emotional responses of participants in a cartographic experiment about interactions with maps, using sensors that monitor psycho-physiological reactions and eye movement data.

3. Experiment

We designed a three-phase experiment, where all subjects participated in all three phases. In the first phase, we study the movement of the participants and their interactions with two different visualizations. We wanted to measure the amount of time that each participant spent looking at each of the posters. Measuring the time participants spend on an activity is a common way of measuring engagement of the activity [LPFK13], including assessing engagement while

Datasets	# Nodes	# Links	# Clusters	Phase
TVCG	2588	3700	34	first
Book	50	75	7	second
Universities	50	125	6	second
LastFM	100	150	4	second
GD	100	250	8	second
World-Trade	200	300	7	second
Ingredients	200	500	4	second

Table 1: *Characteristics of the datasets used in this study. All the datasets are available at:* `http://gmap.cs.arizona.edu/datasets`

working with a visualization [BDF15, HKF15]. We invited each participant to a room where two different types of visualizations (NL and NLG) were installed on two different walls. We then asked the participant to stay in the room for a few minutes until the interviewer came back. The participant was left alone in the room for about three minutes while a camera recorded interaction with the visualizations.

In the second phase, participants were asked to perform three different sets of tasks (node-based, network-based and group-based) using the NL and NLG visualizations with different sizes, densities and datasets. In order to measure the level of enjoyment of performing each specific set of tasks using each visualization, after performing each set of tasks we asked questions about different elements of the enjoyment model [SSK15].

In the third phase, we asked three additional questions. The questions were open-ended in order to allow the participants to convey their feedback and ideas in their own voices and in order to solicit potentially unexpected insights. Specifically, we asked:

- *Having seen the different options, which visualization would the participant select to work with for the remaining tasks?*
- *Why does the participant prefer that visualization?*
- *If the participant can take either of the two posters on the wall with them, which one would they prefer to take?*

3.1. Materials

All relevant materials for this study (datasets, software for running the experiment, anonymized results, and detailed statistical analysis) are available online at `https://bitbucket.org/VIS-HCIResearch/flowevaluationpaper`.

Dataset used in the first phase: The first phase involves evaluating the first impression of the two different visualizations. We selected the **TVCG** dataset to create NL and NLG visualizations. This dataset models the co-authorship network for the IEEE Transactions on Visualization and Computer Graphics, 1995-2015; see Table 1. The vertices represent authors and an edge between two vertices indicates that this pair of authors has co-authored a paper.

Datasets used in the second phase: The second phase involves participants working with different visualizations. To decrease the chance that observed effects are due to a specific dataset, we use six different datasets. In the **Books** dataset nodes are popular books and edges are obtained with a breadth-first traversal following Amazon's "Customers Who Bought This Item Also Bought" links [GHK10a]. In the **Universities** dataset nodes are US universities and edges are constructed based on their similarities in student admissions. In the **LastFM** dataset nodes are popular bands/musicians and edges correspond to similarities between them, as suggested by the online radio station last.fm [GHKV09]. The **GD Collaboration** dataset models the co-authorship network for the Graph Drawing conference [KPS14a]. The **World-Trade** dataset models world trade relationships with edges based on normalized combined import/export between pairs of countries [GHK10a]. The **Recipe-ingredients** dataset contains cooking ingredients extracted from online recipes with edges based on ingredient co-occurrence [AABB11].

Datasets: The nodes in all the datasets are labeled: book names, university names, bands and musicians, author names, country names and cooking ingredients. We selected 200, 100 and 50 nodes from the six datasets and a subset of the links between them to match the desired densities. We also have different settings of 4, 6, 7 and 8 for the number of clusters in the datasets. The clusters are extracted automatically using the GMap implementation of the modularity clustering algorithm [New06]. The graphs are embedded in the plane using multi-dimensional scaling [KW78] as implemented in the "neato" layout algorithm. To generate instances of NLG diagrams we use contiguous GMap [KPS14b]. From the NLG diagrams, we extract the node-link visualizations by removing the group regions. Thus the positions of the nodes and links in the two settings (NL and NLG) are identical. In total we created seven visualizations for each technique (node-link and map-based): the TVCG dataset was used for the posters in the first phase and the rest of the visualizations were used in the second phase; see Table 1.

Size and Density: We use the same settings for size ($N = 50$ nodes as minimum, $N = 200$ nodes as maximum, and $N = 100$ as an intermediate value) and density ($L = 1.5N$ links for the sparse setting and $L = 2.5N$ links for the dense setting) as those in the earlier evaluation of N, NL, and NLG diagrams [SSKB14]. Size is equal to number of nodes and density is equal to number of links divided by number of nodes.

Colors: In the second phase of our experiment the participants are expected to distinguish clusters by correctly naming the different colors used to color the clusters. We use the same map-friendly, qualitative color scheme with eight different colors (red, green, yellow, blue, orange, pink, purple, brown) as Saket et al. [SSKB14]. Specifically, for the visualizations with four clusters we used red, blue, green and yellow. For the visualization with six clusters we added orange and pink. For the visualizations with seven clusters we added purple and for the visualizations with eight clusters we added brown. Node and link colors are also important as the participants need to perform several tasks that assume the readability of the nodes and the links. We use black font for the node labels, RGB (0, 0, 0), and gray links, RGB (90, 90, 90), which were shown to work well in a similar setting [SSKB14].

3.2. Participants and Setting

We recruited 17 participants (10 male, 7 female) aged 21-29 years with normal vision (not color blind). Participants were undergraduate and graduate students with science and engineering backgrounds, familiar with plots, graphs and networks.

The experiment was conducted in a room that was partitioned into two different sections. We ran the first phase in the first half of the room where two posters were installed. The second and third phases were conducted in the other part of the room where a laptop is located. In the last two phases the participants could not see that part of the room where two visualizations were installed. In the second phase, the participants used a laptop (i7-4510U Processor and 14 inch FHD LED Glossy Wedge 1920x1080 screen) and interacted with the mouse to complete a collection of tasks. In the third phase, participants wrote answers to the given questions on a sheet of paper.

3.3. First Phase

In the first phase, in order to capture the user interactions with two posters on the wall, a camera recorded the movement and interactions of the participants with the two visualizations.

Design: We generated two visualizations (one NL and one NLG) and printed them on A1 size paper. The posters showed the 2588-node and 3700-link TVCG dataset with NL and NLG diagrams. Each poster was placed on a different wall of the same room. Since this phase required participants to wait a few minutes while they are alone in the room, we ran a pilot study to determine a reasonable length of time for that unsupervised wait. This is particularly important in our case because if the wait is too short we may not be able to capture sufficient interactions with these visualizations while if the wait is too long the participants might become too bored or upset, or even leave the experiment.

In our pilot study we recruited six participants. We asked the first two to stay in the room while waiting for the interviewer and after two minutes the interviewer asked them about the length of the wait. Specifically, the interviewer asked whether they had enough time to look at both visualizations on the walls. Both participants thought that the wait was a bit short. We repeated the same procedure with another two participants but changing the wait from 2 to 4.5 minutes. Both participants thought that the wait was a bit long. The last two participants waited for 3 minutes and considered this a reasonable amount of wait time which also gave them a chance to take a look at both posters on the walls.

Procedure: All participants were told to meet outside the experiment room and knock on the door upon arrival; at this point the interviewer turned on the camera and invited them inside the room. The participants were asked to read and sign the consent form. They were informed that during the study they will be recorded for five minutes, but not exactly when this will take place. The interviewer then told each participant: "I need to bring some equipment before running the experiment. Please stay in the room for few minutes. I will be back soon and we'll start the experiment." He left the room and came back after three minutes. During this time the camera was recording the interaction with the posters. In order to prevent potential bias introduced by the position of the posters on the walls, half of the participants saw the posters in one configuration, and for the other half the position of the posters was swapped.

Hypothesis: We expected that participants would spend more time (on average) with the NLG poster than with the NL poster, during phase one of the experiment. This hypothesis is based on the earlier results which indicate that participants are more curious about

	Section 1 Node-based	Section2 Network-based	Section3 Group-based	Setting
Visualization 1	Task 1	Task 4	Task 7	NL
Visualization 2	Task 2	Task 5	Task 8	NL
Visualization 3	Task 1	Task 4	Task 7	NL
Visualization 4	Task 3	Task 6	Task 9	NL
Visualization 5	Task 2	Task 5	Task 8	NL
Visualization 6	Task 3	Task 6	Task 9	NL
Visualization 1	Task 1	Task 4	Task 7	NLG
Visualization 2	Task 2	Task 5	Task 8	NLG
Visualization 3	Task 1	Task 4	Task 7	NLG
Visualization 4	Task 3	Task 6	Task 9	NLG
Visualization 5	Task 2	Task 5	Task 8	NLG
Visualization 6	Task 3	Task 6	Task 9	NLG
	First Survey	Second Survey	Third Survey	

Table 2: *This table shows the distribution of nine tasks from various categories over different visualizations. In addition, it shows what tasks participants perform in each section of the second phase.*

NLG visualizations [SSKB15]. Thus, our first **null hypothesis H1** is that there will be no difference in the time spent by participants with the NL and NLG visualizations.

Data Analysis: Two experimenters watched the recorded videos twice and coded the total amount of time that each participant spent looking at each of two visualizations on the walls. The fairly long distance between two posters (about 3 meters) and their position made it hard for the participants to look at both visualizations at the same time.

Results: In total the participants spent 21 minutes and 2 seconds looking at the NLG visualization and 14 minutes and 21 seconds for the NL visualization. We found an effect, albeit not a statistically-significant one ($p = 0.14$, with 95% confidence interval of the true effect being $[-57s, 8.5s]$). In traditional null hypothesis significance testing (NHST) phrasing, we did not find enough evidence to reject the null H1 at $p < 0.05$.

3.4. Second Phase

The second phase contains three sections. In each section the participants performed tasks from one specific category (node-based tasks, network-based tasks, group-based tasks). At the end of each section they took a survey to compare the enjoyment of the two different techniques (NL and NLG) for the corresponding category of tasks; see Table 2. The main reason we asked participants to evaluate two visualization techniques for different task categories separately was to study which visualization is more enjoyable for each category of tasks.

- **Node-Based Tasks**: Tasks in this category can be performed by considering only nodes, so that no other information is required. *For example: Given node X, what is its background color?*
- **Network-Based Tasks**: Tasks in this category can be performed by considering only nodes and links. *For example: Find a node with the highest degree.*
- **Group-based Tasks**: Tasks in this category can be performed by considering nodes, links, and groups. *For example: Given a group X, find all groups neighboring group X.*

Tasks: When determining the tasks that participants need to perform, two goals conflict with one another. On the one hand, we

Node-based Tasks	
T1. Given node "X", what is its background color?	
T2. Find all nodes which start with specific alphabet letter in the specific group.	
T3. What is the number of nodes in a specific group?	
Network-based Tasks	
T4. Given nodes X and Y, find the shortest path between them.	
T5. Find the set of nodes adjacent to a given node.	
T6. Find a node with highest degree.	
Group-based Tasks	
T7. Given nodes X and Y, decide whether these two nodes belong to the same group.	
T8. Find the path X—Y—Z; are nodes X and Z in the same group?	
T9. How many clusters are there in this visualization?	

Table 3: *List of tasks used in the evaluation. We follow its previous use in the literature [SSKB15, SSKB14]. A complete task description, along with Brehmer and Munzner [BM13] discussion of the why/what/how questions about the tasks are provided both in supplemental materials and previous work [SSKB15, SSKB14].*

would like the participants to perform higher level tasks that need deeper understanding of both visualization and dataset: this would suggest the more complex the tasks, the better. On the other hand, we would like the participants to be able to perform the tasks in a reasonable amount of time and workload: this suggests we look for simpler tasks, tasks used in previous taxonomies and evaluations. The final selection was based on tasks representing problems commonly encountered when analyzing relational data, tasks from existing graph/network task taxonomies and used in prior controlled experiments, and tasks that cover a full spectrum of different categories.

With these three main considerations in mind, we selected nine of the tasks utilized in previous work [SSKB15, SSKB14], grouped into three categories based on the information required to perform them:

Most of the tasks in the first two categories are listed under "Attribute-Based Tasks" and "Topology-Based Tasks" in the taxonomy of Lee et al. [LPP*06]. The tasks in the third category are "Group-Based Tasks" in the taxonomy of Saket et al. [SSK14]. Task descriptions of the selected tasks, T1 to T9, are provided in Table 3.

Design: We first created six NLG visualizations using the six different datasets with different size and density. From the NLG visualizations, we obtained the six NL visualizations by removing the group regions. The positions of the nodes and links in the two settings (NL and NLG) are identical. We distributed the three tasks from each category (9 tasks in total) uniformly over the different visualizations. In particular, each task is utilized exactly twice in each setting. We used a within-subject design; overall, each participant performed 36 tasks: 2 settings (NL and NLG diagrams) $\times$ 6 tasks $\times$ 3 sections.

There is no single holistic method to measure the elements of enjoyment. Following the study by Sweetser et al. [SW05], we designed a set of six questions; see Table 4. Each question measures one of the elements (e.g., challenge, focus) of the flow model in visualization. After the participants completed the 12 tasks, using both visualization settings (NL and NLG) in each section, we asked them to measure how well each of these visualization settings met these criteria with a rating based on the Likert scale. We chose the Self-Assessment Manikin (SAM) scale for testing the level of enjoyment of each participant. The SAM scale is a widely used method in psychology that has also found applications in visualization [LDG11, HSF*13]; see Table 4 for the SAM setting.

Procedure: Before the controlled experiment, the participants were briefed about the purpose of the study, data, and techniques used. Although all participants were familiar with graphs and networks, we reviewed all relevant technical definitions (e.g., node, links, adjacency, groups, paths). We then asked the participants to complete six training tasks as quickly and accurately as possible. The participants were highly encouraged to ask questions during this stage and we did not record time and accuracy in this stage.

The main experiment consisted of 36 tasks. Each section consists of 12 tasks for both settings (NL and NLG); as shown in Table 2. In each section, the tasks and visualizations were presented in a random order. Software guided the participants through the experiment by providing task instructions. The participants were able to zoom and pan the visualizations on the screen (if needed) and were required to select one of the provided multiple choices. Once the participants completed the tasks in a given section, the interviewer asked them to complete a survey. The participants were allowed to take a break between sections if needed.

Hypothesis: We expected that in phase 2, the participants would experience more enjoyment, as measured by the average reported level of the six given questions. Our second **null hypothesis H2**, then, is that participants will not exhibit different overall self-reported levels of enjoyment.

Data Analysis: We mapped the 5-point Self-Assessment Manikin (SAM) scale to numbers in the range [-2, 2] with unit increments. Thus we could analyze the data quantitatively.

Results: We found statistically significant effects at $p < 10^{-6}$, giving us sufficient evidence to reject the null hypothesis H2: the 95% confidence interval for the true difference in means being $[0.5423788, 0.2419350]$ (in relative terms, the effect we see could range roughly from 5% to 15% of the total scale).

In addition, we analyzed the results by breaking them down with respect to the questions asked. We would like, after all, to understand which specific aspects of enjoyment account for the overall effect. We illustrate these results in Figure 3. In this case, we repeat the tests in H2, now splitting them into specific question groups, and include a Bonferroni correction factor of 2, to account for multiple comparisons. In this setting, we find that there's enough statistical evidence to reject the corresponding null for questions 1, 4, and 6 (and not enough evidence to reject the null for questions 2, 3, and 5).

3.5. Third Phase

Design: The aim of this phase was to directly collect the preferences of the participants and their subjective enjoyment opinions about the

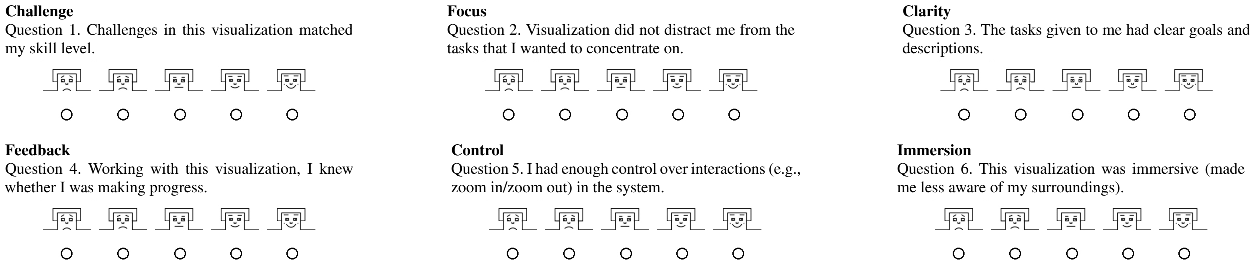

Table 4: *Each survey contains a list of questions designed to measure the elements of enjoyment. We asked participants to select the figure that most closely corresponds to how they feel, following the Self-Assessment Manikin methodology.*

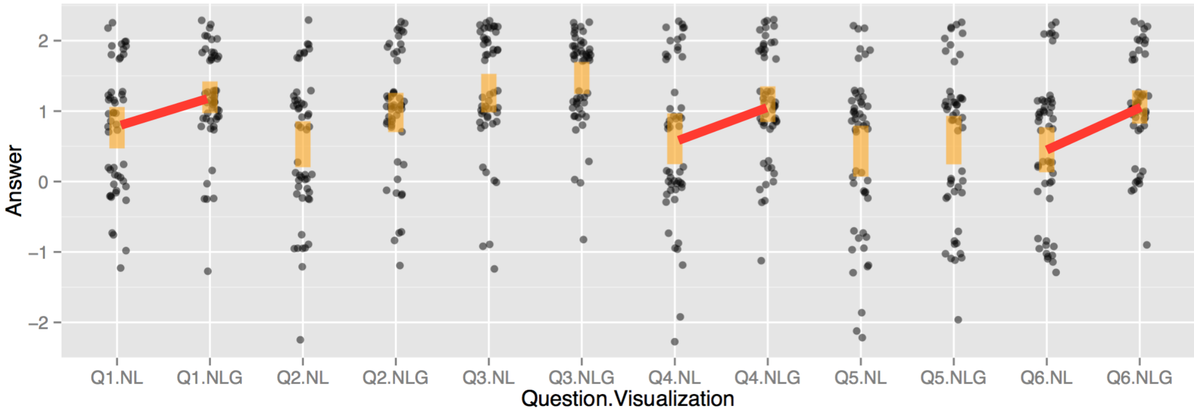

Figure 3: *Summary plot of responses for phase 2. The orange vertical bars indicate the 95% confidence interval for the true overall enjoyment level of each response group. For questions 1, 4, and 6, the difference between the mean enjoyment level is statistically significant at* $p < 0.05$*; we highlight those significant distribution differences by drawing a red line segment between their means.*

two visualization techniques. In this phase we asked the participants to write their answers to three questions.

Procedure: After completing phase one and phase two, we asked the participants to answer the following two questions:

- *If we give you another 6 tasks and ask you to perform them using node-link visualizations or node-link-group visualization, which one would you select to work with?*
- *Why did you select this visualization to work with?* In order to answer this question the participants could select any of options below and they could add others that were not in the list:
  - *because I enjoyed using it*
  - *because it is easier to work with*
  - *because it looks more engaging to me*
  - *because I can perform task faster using it*
  - *because I can perform task more accurately using it*
  - *because it looks more appealing to me*
- *If you were allowed to take one of the posters on the wall with you, which one would you take home?*

Hypotheses: We expected that more of the participants will select node-link-group visualizations to perform the additional tasks. We based this expectation on the previous results, where the participants performed node-based and network-based task with roughly the same accuracy and speed in both NL and NLG settings but both accuracy and time were better with NLG visualizations for group-based tasks [SSKB14]. Our **null hypothesis H3-a** is that the number of participants selecting NL and NLG visualization will be the same.

Further, we expect that more participants would choose to take home the node-link-group visualization. This expectation is based on results in [SSKB15], where participants were more interested and curious about NLG visualizations. Our **null hypothesis H3-b** is that that the number of participants selecting to take home NL and NLG visualization will be the same.

Results: 8 of the 17 participants decided to perform the additional six tasks using the NLG visualization and 9 participants chose the NL visualization. Thus, we could not reject H3-a. The second question in this phase required the participants to answer why they prefer to perform the additional six tasks with one visualization rather than the other. Detailed results for this question are shown in Figure 4. It is worth mentioning that none of 9 participants who chose the NL visualization considered NL visualizations "enjoyable" or "engaging". Moreover, 8 of the 17 participants believe that NL visualizations are easier to use than NLG visualizations. From the 8 participants who chose the NLG visualization:

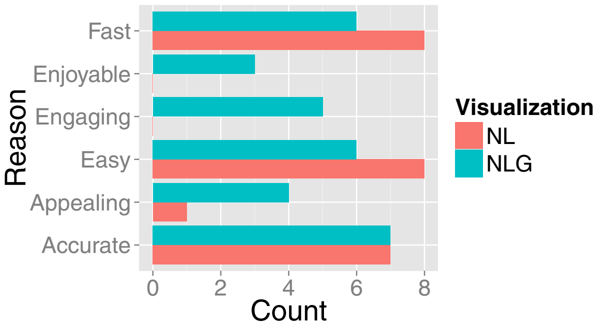

Figure 4: *This figure shows the number of people who selected each specific reason to explain why they preferred one visualization over another. Each participant could select more than one option.*

- 5 of the 8 considered it "engaging"
- 3 of the 8 "enjoyed" using it
- 4 of the 8 found it "appealing"

Results from the last question suggest the appeal of the NLG visualization: 12 of the 17 participants decided to take the NLG visualization home with them. Unfortunately, an exact binomial test indicates that the null hypothesis here ("there's no preference between visualization map types") cannot *quite* be rejected: $p = 0.071$.

4. Discussion

The choice to compare node-link and node-link-group visualizations is motivated by the similarity of the techniques; both are generally applicable to the same kinds of relational data, and this lets us control for more variables. Earlier studies have already examined differences in performance of tasks (in terms of accuracy and time) for these two visualizations, thus making it possible to bring other experimental results [SSKB14, SSKB15] into better context.

From the first phase of our experiment we learned that the total amount of time the participants spent looking at the node-link-group visualization was more than for the node-link visualization; however, the difference was not statistically significant. In the second phase, results indicate that participants believe that node-link-group visualizations are significantly more enjoyable than node-link visualizations. While the first question of the third phase did not yield the result we expected, the second question reveals that even those who decided to work with node-link visualizations did not report that type of visualization as enjoyable or engaging. Results of the third question in the third phase suggest (12 of 17 participants) that more people would decide to take the node-link-group visualization home with them; however, this was not statistically significant.

As discussed in introduction, unlike in previous studies, we hope to understand something beyond just preference of visualization techniques. Specifically, we would like to understand what exactly makes a particular visualization more enjoyable than another? The analysis of the second phase indicates that participants felt that *challenges in the node-link-group visualizations matched their skills* significantly better than in the node-link visualizations. They also felt that it was significantly easier to *follow the progress while working with node-link-group visualizations*. Finally participants believed that node-link-group visualizations were significantly *more immersive* than node-link visualizations. In order to understand which characteristics of the node-link-group visualizations caused participants to give significantly higher scores to these elements (challenges, feedback and immersion), we broke down the responses, based on each category of tasks. The results we report in the following paragraphs should be taken as preliminary, if intriguing, evidence. Further research remains needed.

The participants felt that *challenges in the node-link-group visualizations matched their skills* significantly better only for group-based tasks ($p > 0.5, p > 0.5, p = 0.005$ for Bonferroni-corrected t-tests for respectively, node-based, network-based, and group-based tasks). This shows that the participants felt node-link-group visualizations match their skills significantly better while working with group-based tasks. Based on the previous studies we know that group-based tasks can be performed significantly faster and more accurately using node-link-group visualizations [SSKB15,SSKB14]. Thus we can reject the idea that *difficult tasks would boost the challenges and result in a more enjoyable experience*, because if higher task difficulty is preferred, then (1) participants would not mind a mismatched skill-difficulty setting, and (2) they would consider that the challenges when using node-link visualizations better match their skills.

The participants felt that *following the progress* with node-link-group visualizations was significantly easier while performing group-based tasks ($p > 0.5, p > 0.5, p = 0.002$ for Bonferroni-corrected t-tests for respectively, node-based, network-based, and group-based tasks). One reason could be that the explicit presence of boundaries (closed regions) for the groups in the NLG visualizations more closely matches the notion of a group, and that gives participants more confidence that their answers are correct. Another reason might be that participants felt that following the progress using NLG visualization was easier because they encountered fewer challenges while performing group-based tasks. This could also help participants feel they are accomplishing the tasks. Thus, there might be a relationship between challenge and feedback elements of the flow model. More research is needed here.

The participants felt that node-link-group visualizations are more *immersive*, particularly while performing group-based tasks ($p > 0.5, p > 0.5, p = 0.0021$ for Bonferroni-corrected t-tests for respectively, node-based, network-based, and group-based tasks). In a recent experiment investigating the long-term memorability of underlying data using node-link and node-link-group visualizations of the same data, Saket et al. [SSKB15] found some curious differences between the two visualizations. At the beginning of the experiment the participants were allowed to work and explore the visualizations for as long as they required. On average, participants exploring node-link-group visualizations took 20 seconds longer than the node-link visualizations, roughly a 25% increase. When asked about the reason why, the responses of the participants included "How did you draw this map?","This is beautiful", "I like it". The appearance of the visualization might play an important role in attracting and keeping the attention of the participants. Since it has been shown that eye-movement data can be used to trace cognitive procedures [KDX*12], eye-tracking might help us see differences between the two visualizations. One interesting avenue for future work is to measure enjoyment of other types of visualization techniques to see whether we can find parameters which affect enjoyment of a visualization (*e.g., does using more color in the visual representation make the user's experience with the visualization more enjoyable?*).

There is evidence that the visual appeal of an interface is a factor for enjoyment and engagement [Nor05, LT04]. This relation between aesthetics and enjoyment could be applicable to visualizations. That is, the preferences for NLG visualizations could be partially explained because participants found this type of visualizations aesthetically appealing. NLG visualizations could take advantage of the Gestalt principle of closure, by explicitly enclosing groups of nodes in contiguous regions. It would be worthwhile to further investigate the impact of aesthetic factors on enjoyment of visualization in general, and NL and NLG diagrams in particular.

5. Limitations

We combined three different qualitative (second and third phases) and quantitative (first phase) methods to measure enjoyment and engagement in visualizations. We designed the three phases based on well-known methods which are used to measure enjoyment and engagement in different fields. Since performing such evaluations does not follow a standard template, our study has several limitations.

In phase one, we did not measure the time spent by participants in performing given tasks. Instead, we measured the time spent "looking at a visualization" while they could be doing something else. In other words, we did not ask them to look at, or work with the visualizations; we just asked them stay in the room. Thus, as few of our participants did, they could be singing or dancing, instead of looking at the visualizations. While the total amount of time spent on an activity can be used to measure enjoyment of the activity [LPFK13, HKF15], familiarity or unfamiliarity with the different type of visualization might have an impact on this phase of our study. Results based on self-reporting methods can be questioned, depending on how they are collected and/or interpreted. Fortunately, there is good evidence that people are capable of giving numerical or graphical indication of their emotions [PTTVG03].

Finally, our results are based on a limited number of participants. Our results should be interpreted in the context of the specified tasks, number of participants and visualizations used in this experiment. Despite these limitations, we believe that this study offers evidence that different visualizations could be more or less enjoyable and offers possible starting points for future evaluations of enjoyment.

6. Conclusion

We described a three-phase experiment to evaluate the enjoyment of two different techniques for visualization of relational data. The results indicate that participants found node-link-group visualizations to be significantly more enjoyable than the standard node-link NL visualizations. This is based on the statistically significant differences for three (challenge, feedback and immersion) out of the six measured elements of flow.

7. Acknowledgments

We thank the anonymous reviewers and members of the Visualization Lab at the Georgia Institute of Technology for the constructive criticism and suggestions for improvement. We also thank Avi Dubey for his help in conducting the experiment. This work is partially supported by NSF grant III-1513651.

References

[AABB11] AHN Y.-Y., AHNERT S. E., BAGROW J. P., BARABÁSI A.-L.: Flavor network and the principles of food pairing. *Scientific reports 1* (2011). 4

[ARRC11] ALPER B., RICHE N., RAMOS G., CZERWINSKI M.: Design study of linesets, a novel set visualization technique. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec 2011), 2259–2267. 2

[Art96] ARTZ J. M.: Computers and the quality of life: Assessing flow in information systems. *ACM SIGCAS Computers and Society 26*, 3 (1996), 7–12. 3

[BARM*12] BORGO R., ABDUL-RAHMAN A., MOHAMED F., GRANT P. W., REPPA I., FLORIDI L., CHEN M.: An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2759–2768. 1

[BDF15] BOY J., DETIENNE F., FEKETE J.-D.: Storytelling in information visualizations: Does it engage users to explore data? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2015), CHI '15, ACM, pp. 1449–1458. 3, 4

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. 1, 6

[BMG*10] BATEMAN S., MANDRYK R. L., GUTWIN C., GENEST A., MCDINE D., BROOKS C.: Useful Junk? The effects of visual embellishment on comprehension and memorability of charts. In *CHI '10* (2010). 2, 3

[BVB*13] BORKIN M. A., VO A. A., BYLINSKII Z., ISOLA P., SUNKAVALLI S., OLIVA A., PFISTER H.: What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2306–2315. 1

[CKE11] CERNEA D., KERREN A., EBERT A.: Detecting insight and emotion in visualization applications with a commercial EEG headset. In *SIGRAD 2011 Conference on Evaluations of Graphics and Visualization-Efficiency, Usefulness, Accessibility, Usability,(Stockholm, Sweden)* (2011), pp. 53–60. 3

[CPC09] COLLINS C., PENN G., CARPENDALE S.: Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1009–1016. 2

[Csi90] CSIKSZENTMIHALYI M.: *Flow: The Psychology of Optimal Experience*. Harper Perennia, New York, 1990. 2, 3

[dMTHU10] DE MANZANO Ö., THEORELL T., HARMAT L., ULLÉN F.: The psychophysiology of flow during piano playing. *Emotion 10*, 3 (2010), 301. 3

[EVMJ*11] ELMQVIST N., VANDE MOERE A., JETTER H.-C., CERNEA D., REITERER H., JANKUN-KELLY T. J.: Fluid interaction for information visualization. *Information Visualization 10*, 4 (Oct. 2011), 327–340. 3

[FCPM12] FABRIKANT S. I., CHRISTOPHE S., PAPASTEFANOU G., MAGGI S.: Emotional response to map design aesthetics. In *7th International Conference on Geographical Information Science, Columbus, Ohio* (2012), pp. 18–21. 3

[Fre98] FREDRICKSON B.: What good are positive emotions? *Review of General Psychology 3* (1998), 300–3019. 1

[GHK10a] GANSNER E. R., HU Y., KOBOUROV S.: GMap: Visualizing graphs and clusters as maps. In *IEEE Pacific Visualization Symposium (PacificVis)* (2010), IEEE, pp. 201–208. 4

[GHK*10b] GANSNER E. R., HU Y., KOBOUROV S. G., ET AL.: Visualizing graphs and clusters as maps. *IEEE Computer Graphics and Applications*, 6 (2010), 54–66. 2

[GHKV09] GANSNER E., HU Y., KOBOUROV S., VOLINSKY C.: Putting recommendations on the map: Visualizing clusters and relations. In *Proceedings of the Third ACM Conference on Recommender Systems* (2009), RecSys '09, pp. 345–348. 4

[HKF15] HAROZ S., KOSARA R., FRANCONERI S. L.: Isotype visualization–working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), ACM, pp. 1191–1200. 1, 3, 4, 9

[HR08] HAM F. V., ROGOWITZ B. E.: Perceptual organization in user-generated graph layouts. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1333–1339. 2

[HSF*13] HARRISON L., SKAU D., FRANCONERI S., LU A., CHANG R.: Influencing visual judgment through affective priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 2949–2958. 2, 6

[Jac96] JACKSON S. A.: Toward a conceptual understanding of the flow experience in elite athletes. *Research quarterly for exercise and sport 67*, 1 (1996), 76–90. 3

[Jen00] JENNINGS M.: Theory and models for creating engaging and immersive ecommerce websites. In *Proceedings of the 2000 ACM SIGCPR Conference on Computer Personnel Research* (New York, NY, USA, 2000), ACM, pp. 77–85. 3

[JRHT14] JIANU R., RUSU A., HU Y., TAGGART D.: How to display group information on node-link diagrams: an evaluation. *IEEE Transactions on Visualization and Computer Graphics 20*, 11 (2014), 1530–1541. 2

[KDX*12] KIM S.-H., DONG Z., XIAN H., UPATISING B., YI J. S.: Does an eye tracker tell the truth about visualizations?: findings while investigating visualizations for decision making. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2421–2430. 9

[KPS14a] KOBOUROV S. G., PUPYREV S., SAKET B.: Are crossings important for drawing large graphs? In *Graph Drawing* (2014), Springer, pp. 234–245. 4

[KPS14b] KOBOUROV S. G., PUPYREV S., SIMONETTO P.: Visualizing graphs as maps with contiguous regions. In *EuroVis* (2014), The Eurographics Association (Short Paper). 4

[KW78] KRUSKAL J. B., WISH M.: *Multidimensional Scaling*. Sage Press, 1978. 4

[LDG11] LEWIS S., DONTCHEVA M., GERBER E.: Affective computational priming and creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 735–744. 6

[LM14] LI H., MOACDIEH N.: Is "chart junk" useful? An extended examination of visual embellishment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 58*, 1 (2014), 1516–1520. 2, 3

[LPFK13] LOMAS D., PATEL K., FORLIZZI J. L., KOEDINGER K. R.: Optimizing challenge in an educational game using large-scale design experiments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 89–98. 3, 9

[LPP*06] LEE B., PLAISANT C., PARR C., FEKETE J.-D., HENRY N.: Task taxonomy for graph visualization. In *Workshop on Beyond Time and Errors: Novel Evaluation Methods For information Visualization (BELIV)* (2006), pp. 81–85. 6

[LT04] LAVIE T., TRACTINSKY N.: Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies 60*, 3 (2004), 269–298. 9

[MKK15] MAHYAR N., KIM S.-H., KWON BUM C.: Towards a taxonomy for evaluating user engagement in information visualization. *Workshop on Personal Visualization: Exploring Everyday Life* (2015). 3

[MRS*13] MEULEMANS W., RICHE N. H., SPECKMANN B., ALPER B., DWYER T.: Kelpfusion: A hybrid set visualization technique. *IEEE Transactions on Visualization and Computer Graphics 19*, 11 (2013), 1846–1858. 2

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928. 2, 3

[New06] NEWMAN M. E. J.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA 103* (2006), 8577–8582. 4

[Nor05] NORMAN D. A.: *Emotional design: Why we love (or hate) everyday things*. Basic books, 2005. 9

[OK14] OLSON J. S., KELLOGG W. A.: *Ways of Knowing in HCI*. Springer, 2014. 2

[PA04] PACHET F., ADDESSI A. R.: When children reflect on their own playing style: Experiments with continuator and children. *Comput. Entertain. 2*, 1 (2004), 14–14. 3

[PTTVG03] PAAS F., TUOVINEN J. E., TABBERS H., VAN GERVEN P. W.: Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist 38*, 1 (2003), 63–71. 9

[PYO*13] PECK E. M. M., YUKSEL B. F., OTTLEY A., JACOB R. J., CHANG R.: Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 473–482. 3

[RC05] RATHUNDE K., CSIKSZENTMIHALYI M.: Middle school students' motivation and quality of experience: A comparison of montessori and traditional school environments. *American Journal of Education 111* (2005), 341–371. 3

[SSK14] SAKET B., SIMONETTO P., KOBOUROV S.: Group-level graph visualization taxonomy. In *EuroVis* (2014), The Eurographics Association (Short Paper). 6

[SSK15] SAKET B., SCHEIDEGGER C., KOBOUROV S.: Towards understanding enjoyment and flow in information visualization. In *EuroVis* (2015), The Eurographics Association (Short Paper). 1, 2, 3, 4

[SSKB14] SAKET B., SIMONETTO P., KOBOUROV S., BÖRNER K.: Node, node-link, and node-link-group diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 2231–2240. 3, 4, 6, 7, 8

[SSKB15] SAKET B., SCHEIDEGGER C., KOBOUROV S. G., BÖRNER K.: Map-based visualizations increase recall accuracy of data. *Computer Graphics Forum 34*, 3 (2015), 441–450. 3, 5, 6, 7, 8

[SW05] SWEETSER P., WYETH P.: Gameflow: A model for evaluating player enjoyment in games. *Comput. Entertain. 3*, 3 (2005), 3–3. 6

[TM15] TANAHASHI Y., MA K.-L.: Stock lamp: An engagement-versatile visualization design. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), ACM, pp. 595–604. 3

[VCS02] VASS M., CARROLL J. M., SHAFFER C. A.: Supporting creativity in problem solving environments. In *Proceedings of the 4th Conference on Creativity & Cognition* (New York, NY, USA, 2002), ACM, pp. 31–37. 3