This is an author produced version of a paper published in:
*Information Sciences*

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa20860

**Paper:**

http://dx.doi.org/10.1016/j.ins.2015.04.017

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.

http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/
Can Animation Support the Visualization of Dynamic Graphs?

Daniel Archambault\textsuperscript{a,∗}, Helen C. Purchase\textsuperscript{b}

\textsuperscript{a}Swansea University, Department of Computer Science, United Kingdom
\textsuperscript{b}University of Glasgow, School of Computer Science, United Kingdom

Abstract

Animation and small multiples are methods for visualizing dynamically evolving graphs. Animations present an interactive movie of the data where positions of nodes are smoothly interpolated as the graph evolves. Nodes fade in/out as they are added/removed from the data set. Small multiples presents the data like a comic book with the graph at various states in separate windows. The user scans these windows to see how the data evolves. In a recent experiment, drawing stability (known more widely as the "mental map") was shown to help users follow specific nodes or long paths in dynamically evolving data. However, no significant difference between animation and small multiples presentations was found. In this paper, we look at data where the nodes in the graph have low drawing stability and analyze it with new error metrics: measuring how close the given answer is from the correct answer on a continuous scale. We find evidence that when the stability of the drawing is low and important nodes in the task cannot be highlighted throughout the time series, animation can improve task performance when compared to the use of small multiples.

Keywords: Graph Drawing, Dynamic Graphs, Mental Map, Drawing Stability, Animation, Small Multiples

1. Introduction

Dynamic graph drawing deals with the depiction of graphs that evolve over time in terms of their structure. Nodes and edges can be added and removed. Dynamic graphs are usually represented using a series of \textit{timeslices} or \textit{time steps} whereby snapshots of the graph are taken over given time intervals. For graphs with this representation, dynamic graph drawing algorithms assign coordinates to the nodes in the timeslices in order to depict the structural evolution of the graph over time. Dynamic graphs exist in many application areas including: software engineering \cite{17}, computer networks \cite{12}, distributed systems \cite{47}, sociology and social networks \cite{57, 14}, systems biology \cite{9}, financial networks \cite{58} and many others. Understanding the perceptual factors related to visualizing dynamic graphs is important for these fields.

An important factor to consider when visualizing a dynamic graph is how the information is presented. Animation and small multiples are the two most common, basic methods for the
visualization of dynamic graphs. An animation of the dynamic graph presents the data like an interactive movie whereby the user has complete control of the animation through the use of a slider and a play/pause button. Smooth linear interpolations transition node positions between timeslices and nodes and edges are faded in/faded out if they are inserted/removed from the data. A small multiples presentation of the graph depicts each timeslice of the data in its own window [55]. The windows are arranged in chronological order like a comic book to depict the evolving data. This presentation method has been shown to be more effective than an animation for dynamic graphs on a variety of tasks [6, 25]. More specifically, small multiples is faster with no statistically significant difference in terms of error rate when linked highlighting between timeslices is used.

Animations have the advantage that the full screen can be devoted to a single timeslice at a given time. However, interaction is required and the user must rely on memory to understand how the graph evolves. Small multiples has the advantage that all timeslices are visible on the screen simultaneously and the user does not need to rely on memory. However, the number of pixels devoted to each timeslice is much less and could potentially make the graph more difficult to read.

In dynamic graph drawing, there is always a compromise between the drawing quality of the graph in each timeslice and the stability of the node positions in the drawing as the graph evolves through time. Preserving the mental map [23, 41], or drawing stability, is a property of the dynamic graph that does not allow nodes to move too far in the plane between consecutive timeslices when a change is made to the graph. Thus, stable parts of the graph structure remain in similar positions in the plane during graph evolution. Recent experiments have shown that drawing stability can help with tasks that involve determining the insertion order of specific nodes [29], revisitation to specific nodes in a graph at later times [3], and following long paths through the data [3].

For tasks where drawing stability is important and the positions of nodes remain relatively stable throughout graph evolution, animation can provide little benefit as the identity of each node is disambiguated through its position. However, if the nodes of the drawing move substantially as the graph evolves, animated transitions might provide a larger benefit when compared to small multiples. Situations where the structure of the graph changes substantially due to node insertions/deletions can be the cause of such movement. In many areas of the graph visualization and information visualization literature [10, 46, 8], animated transitions have been shown to be helpful when there are significant changes to the data and/or how it is arranged on the screen.

Tversky et al. [56], strong critics of the usefulness of animation for many tasks in several disciplines, admit that animated transitions can be useful for very specific purposes. Specifically:

- “At this point then, the most promising uses of animation seem to be to convey real-time changes and reorientations in time and space.” [56].

In terms of dynamic graph drawing, animated transitions have been found to be useful when compared to no transitions (like flipping through the slides of a Powerpoint presentation) [10, 8]. Animated transitions have also been explored in the context of encoding changes with colour [5, 46, 8]. However, when comparing animation to small multiples directly on tasks where drawing stability has been shown to help, the question as to whether or not animated transitions provide a benefit when drawing stability is low remains unanswered.

We have previously shown that drawing stability has a positive effect on complex task performance, for both animation and small multiples [3]. In this paper, we investigate a question arising from this result:
For diagrams which have low drawing stability, can animation provide better support for complex task performance than small multiples?

In addressing this question, we analyze the data for dynamic graphs with low drawing stability—graphs with a high degree of node movement between timeslices. We consider task performance not only in terms of response time and error rate, but we also measure the distance between the answer entered by the participant and the correct solution. We therefore apply the following error metrics: euclidean distance, graph theoretic distance, piecewise angle error, and geodesic error. These error measures allow for a continuous measure of error severity rather than discrete correctness values. We find that animation, when compared to small multiples, on low stability drawings can reduce error rates as measured by these metrics.

2. Definitions

In this section, we define what is meant by animation, small multiples, and the mental map. Most of these definitions are consistent across a wide range of experiments that test these factors for a variety of tasks [4].

In this paper, unless otherwise specified, animation means interactive animation. Empirical evaluations have been run where the animations were not interactive, and participants were shown the animation a set number of times [31]. For the experimental data that is analyzed in this paper, participants could interactively view the evolving graph series over time using a slider to see how it changed. Small multiples [55] presents each timeslice in its own window, and the participant reads the timeslices left to right and top to bottom to see the evolution of the dynamic graph. As small multiples showed the entire graph series at once, no interaction was necessary for this presentation method.

For the remainder of this paper, preserving the mental map [23, 41] is referred to using the term drawing stability. Preserving the mental map is a common term used throughout the dynamic graph drawing literature. As the mental map or cognitive map in psychology corresponds to the internal representation of the data inside the human mind, we advocate the use of the term drawing stability for this concept to avoid confusion between our disciplines. Drawing stability excludes the perception of the dynamic graph by the human and refers solely to the positions of the nodes on the external representation of the dynamic graph (usually the computer screen).

We use the definition of Coleman and Parker [20]:

- The placement of existing nodes and edges should change as little as possible when a change is made to the graph.

Many dynamic graph drawing approaches use this definition of drawing stability as a basis for algorithm design [15, 22, 24, 26, 13, 27, 12, 14].

3. Related Work

The experiment and its analysis draw on research from cognitive psychology, perception, and visualization. In this section, we present this related work.
3.1. Principles in Psychology and Perception

Numerous experiments have been run in the field of psychology that provide insight into the factors that influence the usefulness of drawing stability. In the first section, we discuss cognitive maps – how humans understand physical and information spaces. In the second section, we discuss results in parallel object tracking that are relevant to our research.

3.1.1. Cognitive Maps

The term cognitive maps has been studied in a variety of fields with a number of terms for the same concept. Kitchin [37] provides a survey of the area and lists about twenty terms for cognitive maps, including: environmental images, mental maps, cognitive images, mental images, and many others. Tolman [54] was one of the first researchers to use this term to describe how rats, and by analogy humans, navigate their physical environment. Maps are externalizations of this information that help support the cognitive map of the human.

In terms of cognitive maps of diagrammatic representations, Johnson et al. [36] proposed such models existed for logical propositions in order to facilitate explanation and reasoning. Shepard et al. [50] demonstrated that the internal representations of these diagrams were images rather than logical models. However, we cannot fully know the correspondence between the external and internal representations. Palmer et al. [42] defines information equivalence as the extent to which the two embody the same information.

With respect to diagrams, the implications of cognitive maps have been studied in the context of reasoning [30, 53], and for the purposes of learning [21, 1] and thinking [11, 16]. In an information visualization and graph drawing context, the environment is abstract and often does not correspond to actual objects or locations in the real world. Cognitive maps of information, in the fields of information visualization and graph drawing, are supported by overviews. Overviews [51, 33] of the data provide an equivalent concept to maps in this area – a lay of the land or map of the information landscape being explored.

3.1.2. Parallel Object Tracking

When visualizing a dynamic graph using an animation, frequently smooth linear interpolations transition node positions if they change over time. These moving nodes can be viewed as moving points against a field of distractors. There have been a number of experiments on the ability of humans to track objects moving in parallel and the results of these experiments could help us better understand the implications of drawing stability.

Pylyshyn and Storm [44] performed an experiment where participants were to track one to five randomly moving crosses on a screen. At the beginning of this experiment, one to five targets would flash and the remaining crosses would serve as distractors. The entire scene was set into motion where points moved in random directions selected from eight equal divisions of the compass directions. Crosses did not pass too close to each other, but were visually indistinguishable during this phase of the experiment. After seven to fifteen seconds had elapsed, a distractor or a target would be replaced by a box and the participant was asked if the cross was a target. The results of this experiment indicated that participants could track with high accuracy three to five targets. Although accuracy degraded with an increase in the number of targets, 86.5% was achieved even with five targets.

In follow up experiments, Yantis [59] confirmed these results and extended them by demonstrating that coordinated motion of the points increased accuracy further. This result is most likely
explained by the *Gestalt law of common fate* that implies that objects engaged in coordinated motion tend to be perceived together. Liu *et al.* [40] confirmed these results in an air traffic control scenario and extended them by investigating object speed and camera motion conditions. The experiment provided evidence that scene coordinates (the position of objects in the two or three dimensional virtual space) were used instead of image coordinates (the position of those objects on the screen) to perform the task.

As nodes can be viewed as moving points when a dynamic graph is animated, these experiments have relevance to the investigation of the benefits of drawing stability. Formal experiments that test the usefulness of drawing stability should use a sufficiently large number of targets in order to be complex enough so that the participant cannot solely rely on his or her cognitive map. Secondly, these targets should not be preattentively highlighted, for example by using colour, as this highlighting would disambiguate object identity. A third important factor is object speed [40] with rapidly moving targets and distractors being less accurate than slowly moving ones. However, in dynamic graph drawing, the speed of the animation is under control of the participant through the slider. Thus, for experiments involving interactive animation (like the experiment reported here), this property is controlled by the participant.

There are some important differences between the experiments conducted in the area of parallel object tracking and those in dynamic graph drawing. Firstly, collisions between nodes are not avoided by the standard, algorithmic approaches to dynamic graph drawing in the literature [14]. Therefore, for drawings that have low stability in terms of node positions, collisions between parts of the graph often occur. Secondly, coordinated motion is an important factor in dynamic graph drawing. Paths in the graph will exhibit coordinated motion as their nodes are connected by chains of edges. These two factors should be considered when analyzing experimental results.

### 3.2. Dynamic Data

A number of empirical studies have been conducted in the field of information visualization on animated transitions and dynamic data. Heer *et al.* [32] find that animated transitions, specifically staged transitions, can help track changing values in animated statistical graphics. In their experiment, a variety of animated transitions were tested for bar charts, pie charts, and scatter plots. Robertson *et al.* [45] evaluates presentation methods for dynamically evolving scatterplots. The experiment tests representations similar to those of Gapminder\(^1\) – a system for visualizing statistics about the development of various world countries. Tasks either involved an explanation of something in the data to the participant or using the visualization to discover something in the data. Animation, trace lines, and small multiples were tested in this study. Small multiples was found to be the most effective presentation method for all of the tasks studied. Chevalier *et al.* [18] describe an experiment to test animation as a method for highlighting changes in text. In this experiment, the authors found that animation was beneficial in helping illustrate these changes to the document when compared to slideshow – similar to a Powerpoint slideshow that does not use animated transitions.

### 3.3. Dynamic Graphs

Dynamic graphs have received much attention in the empirical studies literature both in terms of animated transitions as well as testing small multiples and animated representations. Empirical studies about both these topics are described in this section.

\(^1\)www.gapminder.org
3.3.1. Animated Transitions

Bederson et al. [10] tested whether or not animated transitions could help with understanding a tree representation of a hierarchy when interacting with it. The experiment found that animated transitions can help in recalling parts of a tree drawing from memory. Shanmugasundaram et al. [49] tested animated transitions in a pan and zoom setting where a graph could not entirely fit inside a viewport, finding that animation did help in this setting. Shanmugasundaram et al. [48] found that animated transitions could help reduce error rates when commuters were trying to find paths from source to destination stations on metro maps. Archambault et al. [5] found that animated transitions were possibly helpful when determining the number of insertions or deletions that had been made to a graph. In a directed graph drawing setting, Zaman et al. [60] found that the appearance of particular nodes and edges could be helped through animated transitions, especially when the differences were highlighted using colour. Animated transitions have been shown to provide improvements over no transitions [8] and the interaction between animation and encoding differences with colour [46] has also been explored.

In general, animated transitions have shown to be useful quantitatively on a variety of tasks in dynamic graphs, specifically when interacting with a graph. Nearly all of these results compare animated transitions to no transitions (similar to a Powerpoint presentation without transitions) and do not consider small multiples. Archambault et al. [5] have small multiples and animation conditions in their experiment, but do not test tasks where drawing stability is required. In this work, we compare animation and small multiples for tasks where drawing stability is known to help [3].

3.3.2. Animation and Small Multiples

Experiments have also looked at animation and small multiples for dynamic graphs. Archambault et al. [6] test animation and small multiples on a number of standard graph readability tasks. They examine changes in node degree, simultaneous appearance of nodes/edges, growth and shrinking of the graph, and shrinking of shortest path distances. For most tasks and overall, small multiples was significantly faster than animation without influencing error rate. Animation significantly reduced error rates for simultaneous appearance questions, but a positive correlation of error rate and time indicates that fast response times did not lead to more errors. Although highlighting helped participants to identify nodes pertinent to the question easily, no other form of visual highlighting assisted them in this question. Farrugia and Quigley [25] tested animation and small multiples in a social network setting. The study found that small multiples was significantly faster than animation for all tasks with no significant difference in terms of error rate.

These studies focused on readability tasks for dynamic graphs and not specifically tasks where drawing stability was known to provide a benefit. Also, key nodes in the questions were usually highlighted using colour through the full time series, causing animated transitions to have less of an effect: if the node pertinent to the question was red, it did not matter where it appeared in the display as it could be retrieved easily by its colour. Experiments [3, 29] have demonstrated that for certain tasks drawing stability is beneficial, suggesting that these tasks are sufficiently complex that they require the use of a stable external representation for understanding the graph. What these results do not tell us, however, is how these complex tasks can be supported as drawing stability degrades. This paper considers this situation, investigating whether the method of presentation (animation or small multiples) might affect the ease with which these complex tasks are performed in dynamic graphs with low drawing stability.
3.4. Summary of Previous Results

Many experiments measure the effectiveness of animation, small multiples, and drawing stability for general dynamic graphs. In this section, we present a short summary of the key results found across some of these studies.

- **Animated Transitions.** Animated transitions have been shown to help when compared to no transitions for many types of dynamic graphs and tasks [10, 48, 8, 46]. These experiments compare animated transitions to no transitions for visualizations where the current timeslice take up the full screen. The no transitions condition is similar to flipping through the slides of a Powerpoint slide presentation. For these cases, animated transitions can help.

- **When linked highlighting is used, small multiples is better and a stable drawing provides little benefit.** If the nodes pertinent to the question are highlighted throughout the full time series of the graph, a more stable drawing provides little benefit [6]. Also, for this case small multiples is consistently faster than animation with no difference in error rate [6, 25]. For an interactive animation condition, this would mean that the pertinent nodes are highlighted through the full time series in a way that makes them visually distinguishable from all other graph elements. In a small multiples condition, this corresponds to linked highlighting across timeslices. If the nodes pertinent to the question are visually distinguishable from all other nodes using colour, stability in terms of node position provides little benefit as the nodes can be retrieved visually by their colour.

- **When does drawing stability help?** If the pertinent nodes cannot be highlighted throughout the full time series making them visually distinguishable, drawing stability has a positive effect for some tasks [3, 29]. These tasks include reconstructing the insertion/deletion order of particular vertices in animated visualizations from memory [29], following long paths through a graph [3], and revisiting one of many specific nodes in a graph [3]. These tasks, in some sense, use the dynamic drawing of the graph as a dynamic map for location of graph elements. In the experiments described above, animation and small multiples were not compared or no difference was found.

In this paper, we look at low stability drawings only and compare animation and small multiples to see if animation can help under these circumstances.

4. Drawing Stability and User Orientation in Dynamic Graphs

We have found that for two complex tasks (revisiting specific locations and following long, specific paths as the graph evolves) drawing stability helps with graph interpretation [3]. In this paper, we consider only the low stability condition of Archambault et al. [3] in order to determine whether animation can provide better support than small multiples for these tasks.

In this section, we describe the experiment and its factors. **Target level** refers to the number of nodes of interest in the question posed to the participant and corresponds to the number of individual nodes or the length of the path tested. These target levels were selected based on the results of previous experiments in psychology as described in section 3.1.2. Experiments [44, 59, 28] further data analysis can be found in Archambault et al. [3].
have provided evidence that humans can track up to five targets in parallel against a field of distractors with both targets and distractors having the same visual encoding. Performance improves if the motion is coordinated. As is appropriate for drawings with low stability, the dynamic graphs were drawn by first laying out each of their constituent timeslices independently using a force-directed algorithm, creating a set of graph drawings which can be viewed as a time series. For these low stability drawings, no information about node positions in one timeslice was used to influence their positions in the next.

4.1. Interfaces

Animation and small multiples were the presentation methods tested in this experiment. In the animation condition, the interface was similar to a movie player (Figure 1(a)). Participants had complete control over the animation and could pause/play it using the button to the lower right or advance the animation at their own rate using the slider. However, no other form of interaction, including zooming, was allowed. The animation started automatically playing after a three second delay if the participant did not interact\(^3\). Smooth, linear interpolations transitioned node positions as the graph evolved. Nodes and edges that were inserted/deleted from the graph were faded in/out of the display. No form of staged transitions were used in this experiment. As only interactive animation was tested, the animation progressed at the desired rate of the participant, controlled directly by the slider. The current state of the dynamic graph series took up the entire screen.

The small multiples presentation method (Figure 1(b)) displayed all six timeslices of the time series as a matrix of images. Participants read this visualization left to right top to bottom to determine graph evolution. During training, participants were instructed to read the matrix in this way. For each individual window, no other form of interaction was allowed, including zooming.

4.2. Data Sets

Two subsequences of the Threads2 [27] dynamic graph were used. The graph encodes online newsgroup discussions and their evolution over time. Nodes are users and edges are replies to posts. Both subsequences were six timeslices long: timeslices 5 - 10 and timeslices 14 - 19 in the original data set. The first subsequence ranged from 46 nodes and 52 edges to 62 nodes and 72 edges. The second subsequence ranged from 72 nodes and 87 edges to 85 nodes and 104 edges. The maximum of number of additions/deletions between timeslices was five for nodes and four for edges. Graph size and series length were selected based on piloting and the experience of earlier experiments [43, 6, 2, 3].

4.3. Tasks

The tasks that were tested in the experiment required participants to exploit the map properties of the dynamically evolving drawing. The first of these tasks was essentially a revisitation task [52, 28], but in a dynamic graph drawing context. The second task required participants follow long paths through the dynamically evolving data. In both cases, the tasks had two phases. In phase one, the participant tracked graph elements through the evolving data and in phase two they entered a response.

\(^3\)This delay was determined by piloting and is consistent with previous experiments [6, 2, 3], allowing participants sufficient time to read the question. It also provided an option to answer the question without interaction with the animation. However, no participant took this option – all of them manipulated the slider directly to progress the animation at his/her own desired rate.
Figure 1: Animation and small multiples interfaces. (a) Animation. (b) Small multiples. The participant used the interface to determine the location of the coloured nodes or the path in the final timeslice. Animation was like a movie player and small multiples is read left to right and top to bottom. Once the participant clicked on Enter Phase 2, the last timeslice of the graph sequence was shown full screen and the participant clicked on nodes to indicate their answer. The time warning label is shown in the small multiples interface.
For the first phase of the task, the participants used the visualization method to locate the individual nodes or the path in the final timeslice of the visualization. When the participant was confident that he or she had found the locations or path in the final timeslice, the participant clicked on Enter Second Phase. In phase two, the final timeslice was shown full screen to the participant and the participant could not return to the previous visualization. The response to the question was recorded by clicking on a single node or the nodes along the path in the final timeslice. The response time reported was that of the sum of the times spent in phase one and phase two. Questions were selected such that no node or edge involved in the question was removed from the data set during evolution.

The location task tested the ability of the participant to revisit a specific node in the graph. In a more maplike setting, this question is similar to locating landmarks in a city (e.g. Where is St. Stephen’s Green in Dublin?). The targets were spread through the graph and were independently moving nodes, initially coloured in isoluminant shades of blue, red, yellow, green, or purple. All other nodes were shaded a light grey. Targets were only highlighted with colour in the first timeslice and were not highlighted in subsequent timeslices. After the first timeslice, the colours of the targets were set the same light grey colour as all other nodes in the experiment. The number of coloured nodes defined the target level for the question: one target for low, three targets for medium, and five targets for high. The participant used the visualization to determine the location of all nodes in the final timeslice.

In phase 1 of this question, participants were asked:

• Please determine the location of all targets in the final timeslice.

Subsequently, in phase 2, participants were asked:

• Where is the <colour> vertex?

One of the colours listed above was the value for <colour> in the question.

The path task tested the ability of the participant to follow a long, directed path through the evolving graph. In a more maplike setting, this question would be similar to finding a route through a city following streets and intersections (e.g. What streets do I take to get from St. Stephen’s Green to O’Connell Bridge?). The targets corresponded to individual nodes along the path and were initially indicated with a blue colour. Saturation indicated the direction of the path and participants were instructed to enter the path from darkest blue to lightest blue in the second phase. Nodes were highlighted in the first timeslice but not in subsequent timeslices. After the first timeslice, these nodes faded to the same light grey colour as all other nodes used in the experiment. As paths are connected, the motion of these targets was coordinated and a larger number of nodes were tested. Informed by Yantis [59], the number of targets (the length of the path) is 3, 5, and 7 nodes for low, medium, and high target levels respectively. The participant used the visualization to determine the location of all nodes in the final timeslice.

In phase 1, participants were asked:

• Please determine the path in order from the darkest vertex to the lightest vertex in the final timeslice.

In phase 2, participants were asked:

• Click on the vertices of the path in order (X vertices).

The value of X was the number of vertices on the path.
4.4. Experiment Design

The experiment was divided into two, counterbalanced blocks (animation and small multiples) of 3 target levels × 2 data sets (subsequences of Threads2) × 2 questions, resulting in twelve tasks for each of the two presentation methods under investigation.

The tasks were randomized individually for each participant and were prefixed with eight practice tasks. The practice tasks presented each data set once, each question four times, and each target level at least twice. The results of the practice tasks were discarded and participants were not made aware that these tasks did not form part of the experiment. These tasks were divided into two blocks between which participants could take a short break.

Each participant answered all questions under one presentation method first and the other second, reducing any cognitive shift required to move from one interface to another as it occurred only once. Animation was presented first to even-numbered participants while small multiples first to odd-numbered participants. Before the start of each presentation method block, participants had a demonstration session which introduced them to the experimental interface. In these demonstration sessions, participants could ask questions, find out about the experimental tasks and learn how to find the answers to the questions. Participants were also tested in these blocks to ensure that they could distinguish and name the colours used in the experiment.

Both interfaces were rendered in real time using the Tulip framework [7]. No time limit was enforced per task or for the experiment overall. However, a warning label appeared on the screen after fifty seconds had elapsed for each task, and participants were encouraged to finish the task after that point. Overall, there were twenty-eight (24 male, 4 female) participants used in the final results. Participants were drawn from members of the Complex and Adaptive Systems Laboratory of University College Dublin (UCD CASL). On average, each participant took an hour to complete the experiment.

5. Measurements of Error

Archambault et al. [3] found that there were strong effects of drawing stability for both animation and small multiples but no statistically significant differences between them. This was, in part, due to the fact that the effect of drawing stability was so strong for these tasks.

In order to determine if animation has an effect on performance when the dynamic graph drawings have low stability, we analyze our data from a different perspective, using a new series of metrics that measure continuous distance to the correct answer rather than discrete correctness. We assume that answers closer to the intended target are better than answers far from the intended target. These metrics are described below and illustrated in Figure 2 for the location task and Figure 3 for the path task.

5.1. Absolute Error

The absolute error is a measurement of whether or not the participant selected the correct nodes exactly. It does not take into account how close the participant was to the correct solution.

For the location task, the answer was considered correct if the answer entered by the participant was exactly the same node as the correct answer. Thus, correctness was measured as a binary value with answers being correct or incorrect only.

For the path task, correctness was determined using \((X - d)/X\) where \(X\) was the length of the path in the answer and \(d\) was the Levenshtein distance (edit distance) [39] between the nodes of the
entered answer and the correct answer. Edit distance measures the number of insertions, deletions, and changes between the entered answer and the correct answer and is at most $X$ (replacement of the entire string). This measure only considered *exact matches* to the nodes highlighted in the first timeslice of the task.

5.2. Euclidean

The Euclidean error tries to quantify the degree of error in the simplest way through the straight line distance between the entered and correct solution.

The Euclidean metric measures the distance between the answer entered by the participant and the correct answer in the final timeslice as the distance between the Euclidean coordinates of the two nodes (Figure 2(b)).

For the path task, this distance is the average, pairwise distance between corresponding nodes of the directed paths in entered and correct answer (Figure 3(b)).

5.3. Shortest Path

The shortest path error tries to quantify the degree of error using only the connectivity of the node-link diagram independent of its drawing. This distance is the unweighted shortest path distance between the entered and correct solution.

The shortest path metric measures the shortest path distance between the answer entered by the participant and the correct answer as the shortest path between the two nodes in the final timeslice (Figure 2(c)).

For the path task, the measure is the average shortest path between corresponding nodes of the directed paths in entered and correct answer (Figure 3(c)).

5.4. Piecewise Angle

Piecewise angle measures how closely the participants followed the general direction of the path through the graph. It quantifies if they took the right turns at the right time while navigating through the data.

This metric measures piecewise angle deviations between the entered and the correct answers for the path task only (Figure 3(d)). For each pair of nodes along the paths of the entered and correct answers, the angle between the vectors is computed and averaged. As explained previously, the direction of these paths is from darkest to lightest node.
Figure 3: The metrics measured for the path task in the experiment. The direction of the path is always from the most saturated to the least saturated colour. 

(a) The graph with the entered answer highlighted in blue and the correct answer highlighted in green. The direction of the answer is from least saturated node to most saturated node. 

(b) The Euclidean error metric ($E_e$) measures the average Euclidean distance between corresponding nodes in the entered answer and the correct answer to the question. 

(c) The shortest path metric ($E_{sp}$) measures the average shortest path between the entered answer and the correct answer to the question. 

(d) The piecewise angle ($E_{ang}$) metric measures the average piecewise angle between the entered answer and the correct answer to the question. 

(e) The geodesic angle ($E_{geo}$) measures the angles between the geodesics of the entered and correct answers to the question.
5.5. Geodesic Angle

Previous experiments [35] suggest that participants use the straight line between the start and end node of a specific path as an estimate of the path between two nodes. As such, we can use it as an error measure. Also, it measures whether or not the participant was able to get the general direction of the path through the data at a high level.

This metric measures the geodesic error between the entered and correct answers for the path task only (Figure 3(e)). The measure compares the geodesic of the entered path to the correct path as an angle.

6. Results

In this section, we analyze our data with respect to our research question: “For diagrams which have low drawing stability, can animation provide better support for complex task performance than small multiples?” We analyze this data with respect to the metrics defined above. As the location and path tasks are quite different in nature, we treat them separately.

A Shapiro-Wilk test, with a significance level of $\alpha = 0.05$, was used to determine the normality of each of our distributions overall. Usually, it was found that at least one of our distributions was not normal. Thus, we applied an exact Wilcoxon signed rank test when comparing most of these distributions with a significance level of $\alpha = 0.05$. In the few cases where both distributions were normal, we checked the results with a t-test at the same significance level and found that there was no difference in our results for all cases. When the data sets were divided by target level, we further reduced the significance level to $\alpha = 0.017$ and used the same procedure as described above. Usually, one of our distributions was not normal, but in the cases where both distributions were a t-test did not affect our results.

In all bar charts, black lines connect pairs with statistically significant differences and below such differences the p-value is indicated. Mean and median values, separated by a hyphen (-), are indicated below each bar. The standard error is also indicated. The results of this analysis are presented in Figures 4, 5, and 6 respectively. The most important statistically significant results to note are:

- For the **location** task (Figure 4)
  - Animation produced answers closer to the correct answer overall and for one target on all measures
  - For three targets, animation produced closer answers for Euclidean and shortest path metrics
  - There was no evidence to support either presentation for five targets
  - In terms of response time (Figure 5), animation was faster overall and for one target

- For the **path** task (Figure 6)
  - Animation produced answers closer to the correct answer overall and for paths of length three on all measures
  - For paths of length five, animation produced closer answers for the absolute error, average Euclidean distance, and shortest path metrics
Figure 4: Results of the various error measures, as defined in section 5, for the location task.

- For paths of length seven, animation produced closer answers for all measures
- In terms of response time (Figure 5), animation was faster overall and for paths of length three

7. Discussion

In Archambault and Purchase [3], the effect of drawing stability was pronounced and the only differences found between the factors of the experiment were between the low and high stability factors. However, by considering different measures of error and the low stability drawings, we can begin to see effects when there are drastic changes in orientation in the data. For these cases, animated transitions can be more helpful in tracing specific paths and revisitation tasks when compared to small multiples.

For the location task, when using the absolute error measure, we find very slight statistically significant differences overall and for a single target. When considering the new error metrics of Euclidean distance and shortest path, we are able to see much stronger statistically significant
improvements overall and up to three targets for the animation condition when compared to small multiples. In terms of response time, there are also very slight significant differences overall and for up to three targets.

This result provides evidence that animation can improve participant performance, when compared to small multiples, for drawings with low stability. It provides even stronger evidence that animation, under these same conditions, helps the participant answer closer to the target. In previous experiments which have tested animation and small multiples, animation required significantly more time with no difference in terms of error rate. However, in this experiment, all drawings had low stability. Thus, for revisitation tasks on dynamic graph drawings of low stability, animation can help. In some senses, this result complements results which find that animation, when compared to no animated transitions, can help for certain tasks [8, 46].

It is interesting to note that generally the Euclidean distance and shortest path error measures produced the same results throughout our analysis. This result could probably be due to the fact that generally force-directed algorithms try and keep nodes that are close, in a shortest path sense, to each other in the drawing. Similarly, piecewise angle and geodesic produced similar results. As both of these metrics provide a measure for reading the correct direction of a path, it makes sense that the statistically significant results witnessed by these two metrics would be similar.

For the path task, when using the absolute error, we find statistically significant differences overall and for all three path lengths. These results persist for the Euclidean distance and shortest path measures. For the piecewise angle and geodesic errors, we find results overall and for the paths of length three and length seven. In terms of response time, there are very slight significant differences overall and for paths of length three.

This result provides evidence that animation, when compared to small multiples, can improve
<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Length 3</th>
<th>Length 5</th>
<th>Length 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Error</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Euclidean Distance</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Shortest Path</td>
<td><img src="image9" alt="Graph" /></td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Piecewise Angle</td>
<td><img src="image13" alt="Graph" /></td>
<td><img src="image14" alt="Graph" /></td>
<td><img src="image15" alt="Graph" /></td>
<td><img src="image16" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Geodesic</td>
<td><img src="image17" alt="Graph" /></td>
<td><img src="image18" alt="Graph" /></td>
<td><img src="image19" alt="Graph" /></td>
<td><img src="image20" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
</tr>
</tbody>
</table>

Figure 6: Results of the various error measures, as defined in section 5, for the path task.
the ability of participants to follow long paths through the data for drawings of low stability. The
selected nodes are, on average, closer to the intended path and generally the entered path has the
correct direction through the data. Thus, for these path tasks under these conditions, animation
can help. In previous experiments which have tested animation and small multiples, animation
required significantly more time with no difference in terms of error rate. However, this result
provides evidence that animated transitions, when compared to small multiples, on low stability
data can help improve performance. As stated for the location task, this result complements
results which find that animation, when compared to no animated transitions, can help for certain
tasks [8, 46].

For the path task, when considering piecewise angle and geodesic, we found that animation
could help the participant get closer to the correct direction through the data overall and for paths
of length three and seven, but not for five. For our data sets, it is probably the case that our paths
of length five strayed further from the geodesic during evolution, making it more difficult to follow
the direction of the path, but further study is required to answer this question.

It is interesting to speculate on why this experiment shows that animation is useful under the
conditions of low stability, no highlighting using colour, and with continuous error metrics. Our
first observation is a methodological one: we believe that continuous error metrics are more valid
than the binary ones (used in previous experiments) when the task is to identify a location or path –
they do not simply discard incorrect responses, but take into account the extent to which the
response is almost correct. It may indeed be that prior experiments when re-run using continuous
metrics may show further support for animation under low stability conditions. Secondly, low
stability animations result in nodes moving large distances, and animation helps the user follow
nodes as they move. Finally, two of our metrics measure the direction of paths through the data,
a quantity not previously measured, and animation, under low stability conditions, may be able to
help users follow this change in direction.

8. Limitations and Future Work

Our experiment was designed to test animation and small multiples in the case of low drawing
stability on two questions: revisiting nodes in a graph and tracing long specific paths in a
graph. The experiment varied the number of targets using results from the literature. However,
testing a wider range of data set sizes and more questions would have made the experiment pro-
hibitively long in duration. The generalization is therefore limited by these parameters and further
experimentation is needed.

An interesting next step would be to employ eye tracking methodologies [34, 38] to provide
an interesting new perspective on questions relating to the utility of animation, small multiples
and drawing stability. Such studies may help illustrate the root causes of the effects that we have
measured with previous experiments.

Our experiments are conducted on data sets where the number of nodes in the graph is in the
hundreds. Although this is small when compared to real data sets, the approach to deal with larger
graphs often relies on either filtering or aggregation across the nodes of the graph to reduce them
to a reasonable level. In this case, our results apply directly. However, it would also be interesting
to investigate effective techniques and evaluations for visualizing graphs of this size more directly.

When we conducted our experiment, all of our dynamic graph series had six timeslices. This
decision was made to be consistent with previous experiments and make their results comparable.
Thus, our results apply to time series of about this length and real data sets can often be larger.
However, most techniques to visualize dynamic graphs with long time series usually employ either filtering or aggregation across the number of timeslices to reduce this number to a reasonable size. In this case, our results can be applied directly. In future, it would be interesting to investigate techniques and evaluations that are able to visualize long time series more directly.

A number of techniques, as previously mentioned, employ differencing methods and staged animations. Empirical evaluations have been already conducted [5, 8, 46] on such methods as well. In a context where colour is used for the encoding of attributes or other data on the nodes and edges of the graph, differencing methods cannot be used. In this case, it helps to understand graphs without these encodings. As our research question was primarily interested in the difference between animation and small multiples in this context, we chose the simplest form of animation. Therefore, the results here may not be further impacted by these techniques. However, a recent experiment has found that staged transitions for certain tasks do not bring large benefits and can sometimes worsen performance [19]. On a philosophical note, it is important to understand what is going on at a basic level, without extra factors potentially inducing confounds into the experiment, to progress these scientific questions forward.

In this paper, we have only investigated one form of interaction with a dynamic graph: interactive animation. This form of animation is by far the most common and most tested form of interaction in the dynamic graph drawing literature. We would strongly encourage the design and evaluation of other novel and effective interaction methods with dynamic graphs.

9. Conclusion

The conclusions of any experiment are bounded by its parameters. While we have used real-world data, we have only used one data set, and have amended it to our purposes; while we have used complex dynamic tasks previously shown to require support from an internal cognitive map, there are only two tasks; our participants are from a particular demographic with specialised expertise.

Despite these limitations, these results demonstrate for the first time support for animation of dynamic graphs, when compared to small multiples, under given circumstances. This result provides a firm basis on which a range of wider communities, tasks and data can be investigated.

Archambault et al. [3] found that drawing stability was important for disambiguating the identities of specific nodes in a graph and long specific paths. This effect was seen both for animation and small multiples. Further analysis of this data for graph drawings of low stability finds that animation can improve participant performance, when compared to small multiples, on these tasks. This result suggests that if there are many reorientations in a dynamic graph drawing and important nodes in the task cannot be highlighted using colour through the full time series, animation can help the participant better answer these tasks.

We conclude that given a choice between presentation method and drawing stability, drawing stability is likely to provide the largest benefit for these tasks. If the drawings can be rendered with reasonable stability or the pertinent nodes to the task can be highlighted using colour, small multiples often has significant advantages in terms of response time as shown in previous experiments on dynamic graphs [6, 25], experiments in other areas of information visualization [45], and in surveys on the topic [56]. However, if the dynamic graph drawings are not stable and highlighting cannot be used, animation can provide benefits over small multiples.

These results complement those that demonstrate that animation can provide benefits over no transitions for full screen and hybrid visualization techniques [10, 46, 8] and techniques that encode
changes with colour [5, 46, 8]. Also, it confirms the statement of Tversky et al. [56], for dynamic graphs, that animation is useful in conveying real time changes and reorientations in space when many such reorientations occur in the data over short periods of time.

Acknowledgements

This work was supported in part by Clique Strategic Research Cluster funded by Science Foundation Ireland (SFI) Grant No. 08/SRC/I1407.

References