EvolutionWorks

Towards Improved Visualization of Citation Networks

Jason Wilkins[(⊠)], Jaakko Järvi, Ajit Jain, Gaurav Kejriwal,
Andruid Kerne, and Vijay Gumudavelly

Texas A&M University, 3112 TAMU, College Station, TX 77843-3112, USA
jwilkins@tamu.edu

Abstract. *EvolutionWorks* supports exploratory browsing of the academic paper citation network with an animated and zoom-able visualization that helps researchers explore the conceptual space that emerges from the relationships between academic papers. Metaphorically speaking, a researcher starts out with the *seed* of an idea that will *grow* into an unwieldy set of potentially useful papers that the researcher must *prune* into a final reading list. Accordingly, EvolutionWorks provides novel affordances to explore the citation network based on this *seed-grow-prune* model. First, *kinetic layering* represents abstract document properties as physical properties in a force-directed layout. Second, a *unified layout* shows the network graph and documents in a single view. Third, the *focus-context-focus hop* is a way to change focus from paper to paper that keeps researchers aware of the immediate context. Finally, if there is a tight cluster of papers, the system automatically creates *cluster summary titles* that are easier to read.

Keywords: Citation networks · Graph visualization · Information retrieval

1 Introduction

Making sense out of the body of academic literature is a challenge. In this paper, we treat the literature primarily as a *citation network*. The citation network is a large conceptual space embodied in the complex hypertext constructed from all the various connections between academic research papers—such as references, citations, or shared authorship. Although interactive visual presentations of graphs can aid people who need to explore and comprehend complex data [1–3], like the academic literature, only a few experimental search tools currently take advantage of such presentations. More often, open public access catalogs and web-based search tools have user interfaces that unimaginatively present users with pages of text full of hyperlinks. A specific problem is that tools present search results primarily as text, which fails to take advantage of most *retinal variables* [4] (size, color, value, shape, etc.), so the amount of information they can convey about the relationships between research papers is

relatively impoverished. The result is an additional burden placed on the researcher to construct and maintain more of the conceptual space of the academic literature within their own limited working memory. With these issues in mind, we have developed *EvolutionWorks* to help researchers navigate the academic literature effectively by presenting it as an interactive visualization of the citation network that takes advantage of more retinal variables to characterize more strongly the relationships between papers.

The main scenario that guided the design of EvolutionWorks is that of a researcher building a *reading list* for a potentially unfamiliar topic. For our purposes, a reading list is a set of academic research papers, relevant to a particular topic that a researcher saves to consider in more depth after scanning the literature. More broadly, building a reading list is an *information-based ideation task* [5] that requires a researcher to iteratively explore, think about, collect, and develop significant ideas. Further, we define *exploratory browsing* [6] as the creative process that researchers engage in when they seek diverse and novel information as they use the Internet to investigate a conceptual space. In this case, the researcher follows chains of references and citations. We designed the features of EvolutionWorks for the purpose of aiding researchers in exploratory browsing by making it more clear what papers are important and what links to follow-up, with the aim being to paint a clearer picture that grants the researcher greater confidence that their reading list will improve their comprehension of the state of the art.

The contributions of EvolutionWorks that we present here are the *seed-grow-prune design model, kinetic layering,* a *unified presentation, focus-context-focus hopping,* and *cluster title summarization*:

- Our *seed-grow-prune* design model reflects what we discovered from interviewing researchers about how they approach the task of building a reading list. EvolutionWorks supports the metaphorical task of a researcher starting out with the *seed* of an idea that will *grow* into an unwieldy set of potentially useful papers to read that the researcher must *prune* into a final reading list.
- Our *kinetic layering* methodology is the encoding of the abstract properties of documents into the physical properties of the particles that drive a force-directed animation engine that dynamically positions the nodes of a graph visualization. Thus, the physical simulation helps to make complex relationships and properties clear by visual inspection, for example, by causing related papers to group automatically because they attract one another, or by making important papers appear more significant because they are bigger and heavier.
- There are drawbacks to presenting paper content and the citation network in separate views, such as a split screen or by separate paper and network viewing modes. Specifically, a split presentation takes up extra screen real estate (which is scarce on mobile devices), and separate modes create a cognitive disconnect by dividing the researcher's attention between two representations of the same paper. To avoid this, EvolutionWorks combines both papers and the citation network into a single *unified presentation* with papers represented directly as graph nodes.
- The unified presentation alone would have a problematic trade-off between zooming in to view a paper's content and zooming out to keep track of the broader

relationships between papers. We developed a context preserving navigation primitive, the *focus-context-focus* hop, which helps to keep the larger context in mind. When a researcher selects a paper link, the system brings the target paper into focus by smoothly panning from the original view to the subsequent view. In order to give the researcher a sense of the larger context the view also zooms out to reveal more of the citation network before zooming back in to the new focus. The impression given is much like that of physically hopping from paper to paper.

- Another problem with the unified presentation is that zooming out tends to cause paper titles to overlap and become illegible. To solve the problem EvolutionWorks automatically takes tight groups of papers and collapses them into *clusters*, and gives each cluster a single *summary title*, computed based on the metadata of the clustered papers.

To evaluate EvolutionWorks we conducted a small user study. The results were indicative that all in all these contributions work towards creating an improved visualization of citation networks compared to a typical Web browser.

The rest of this paper is organized as follows. Section 2 provides background material that justifies the need for EvolutionWorks and the science that supports its design. Section 3 briefly reviews related work by other researchers in this area. Section 4 explains each contribution of EvolutionWorks in more detail and discusses a few implementation concerns. Section 5 reports the results of a user study that evaluates how well EvolutionWorks steers users towards better papers. Section 6 puts forward possible avenues for further improvement and evaluation of this work and concludes the paper.

2   Background

The Google search engine has become so ubiquitous that the term "googling" and research are seen as one and the same [7]. Whether or not this is a good thing, (no consensus has been reached), users have come to expect search tools to be simple [8]. It seems that simplicity and intuition trump the organization and quality of search results; so much so that many would rather search the World Wide Web for academic papers than use online public access catalogs (OPAC) that have been shown to produce higher quality results [9]. Since users find OPACs less intuitive [8, 9]—and the Web is relatively disorganized—there is a need for a more intuitive search tool specifically tailored for the particular structure of the academic literature [10, 11]. This section reviews the research background that we relied on to build such a tool; one that aims to be as friendly as a browser, but leads users to higher quality search results.

The s*eed-grow-prune* methodology is derived from Bates' *berry picking* model for on-line searches [12] and from the *mixed-initiative* ideation engine *combinFormation* [13]. When describing berry picking, Bates made the case that searches evolve as the user gathers more information and zeroes in on what they need—which may actually end up being quite different from what they originally thought they needed. Additionally, the mixed-initiative system combinFormation combines the actions of the system and the user in a single collaborative space. Both the system and the user can

modify the collaborative space, with the user providing guidance to improve the system's actions.

EvolutionWorks relies on *physically based modeling*, which is a methodology were objects in some model are given physical properties—such as position, velocity, acceleration, and mass—that are acted on by forces. The system simulates the behavior of objects according to the object's given properties, with the goal being that people will recognize the resulting animation as being physically plausible. One case of a physically based simulation is a *force-directed* graph layout, defined as a simulation of the motion of graph nodes as if they were point masses connected through the edges of the graph by springs [14–17]. The system assigns physical properties to the nodes and edges, and the results of the simulation determine node locations. To keep nodes from overlapping, the simulation adds a repulsive force between nodes to push them apart. Interactive force-directed layouts allow the user to tweak the layout while the simulation responds intuitively to the user's intervention.

The *kinetic layering* methodology, build upon force-directed layout, is analogous to the *visual layering* [18] and *kinetic visualization* [19] methodologies. With visual layering, the *retinal variables* [4] of location, size, shape, color, value, texture, and orientation are used to encode information, while with kinetic visualization, animated contours disambiguate 3D surface features. There are several examples of how the perception of motion is an important source of information and is the key to why kinetic layering is useful. First, the *kinetic depth effect* [20] allows the 3D shape of an object to be understood, even if all that is visible is the silhouette of the object, but if and only if the object is moving. More generally, motion allows the brain to segment an image into its separate parts. Second, people are especially sensitive to *biological motion* [21]. People can readily recognize the characteristic motion of plants and animals from relatively sparse information. For instance, a grouping of white dots on a black background might appear completely random if they are static, but it is instantly apparent what they represent if the dots start moving as if attached to a tiger. Certainly, motion that mimics life is compelling. Finally, motion also helps to overcome a certain kind of perceptual glitch caused by the very short-term nature of sensory memory. Under some circumstances, suddenly adding or removing an element from an image might cause an observer to remain oblivious to any change at all. This effect is called *change blindness* [22] and it represents a fundamental limitation in the ability to perceive changes that are not continuous.

3   Related Work

There have been many efforts undertaken to meet the need for improved tools for exploring networks and in particular, the academic citation network.

Vizster [2] is an interactive visualization tool for exploring the community structure of on-line social networks such as *Facebook*. It draws a springy animated *sociogram* (social graph) driven by force-directed layout, and is in fact one of the main inspirations for *EvolutionWork's* design. Vizster's creators clearly demonstrated the advantage of graph visualization in a party setting where people were surprised to discover *friend of a friend* relationships that were not apparent to them through the ordinary web

interface. Although Vizster uses a force-directed layout to create a springy and "fun" visualization, it does not take advantage of physically based modeling to convey any additional information.

There have been several systems built that directly address the problem of exploring the citation network. One such tool is *CiteWiz* [23], which supports multiple visualization techniques. First, the *growing polygon* [24] is used to visualize causality and was found to illustrate partially ordered sets better than Hasse diagrams. Taking advantage of this, CiteWiz uses growing polygons to show the influence of authors on each other as derived from the partial order of the citation network. However, the growing polygon visualization provided by the CiteWiz tool is less intuitive than a Hasse diagram, so it requires some training before users can take full advantage of it. Additionally, its use of color limits accessibility for the colorblind. The second visualization provided by CiteWiz is *Newton's shoulders*, which represents a timeline of when authors started publishing combined with how many citations they have. In the visualization, each stick figure represents an author whose size is determined by that author's total citation count, and later authors are stacked on top of the earlier ones. The final visualization is CiteWiz's *concept map*, which is a force-directed layout that can show the relative importance of either keywords or authors. In the case of keywords, larger nodes represent keywords that are more common and springs connect keywords that appear together. In the case of authors, larger nodes are authors with more citations while springs connect authors that worked on the same paper. Applying force-directed layout to keywords and authors, but not papers, seems like a missed opportunity.

PaperCube [25] is a web application that supports several different visualization methods for both papers and authors, but we will only mention the *CircleView* [26] method here. CircleView shows *focus+context* for a single paper by placing the focus paper in the middle and encircling it with the context—the focus paper's citations. CircleView begins to break down when the number of citations is high. The limit appears to be about 15 citations, but papers commonly have many more. The context provided by CircleView is also limited to the immediate relationships between papers. The system does not smoothly animate the transitions between focused papers, which can cause problems due to change blindness.

A *co-citation* occurs when two separate papers both cite another paper. The purpose of *CociteSeer* [27–29] is to visualize the relationships between papers that co-citation implies. Rather than explicit edges between co-citing papers, CociteSeer implies these relationships by placing co-citing papers close to one another.

Physically based modeling has been used to visualize academic *career histories* [30] derived from citation data. Two different visualizations were reported: the first modeled an author's publication patterns as fluid flow through pipes and the second modeled an author's influence as wave propagation through a graph.

A *research front* [31] is a set of documents that cite a base set of fundamental documents that is fixed and time invariant. By visualizing research fronts, it is possible to determine how research splits into different branches over time.

PaperViz [32] makes use of the radial filling space and bulls-eye view visualization techniques that attempt to show the relationships between papers while saving screen space.

The *XML3D* [33] web browser demonstrated the potential of novel visualization methods to enhance the user experience. The browser has an interactive hyperbolic 3D graph view of pages and links. The hyperbolic view is a focus+context visualization that keeps global relationships between documents visible to the user. Promisingly, a user study found that the browser increased both user productivity and satisfaction, which corroborates with our results. However, to show the full benefit, the XML3D web browser had to provide the user with two separate views: a hierarchical summary of links, and the hyperbolic view. Neither the link summary nor the hyperbolic view alone showed a clear advantage. EvolutionWorks integrates both views into a single unified presentation.

4   EvolutionWorks

This section describes the features of EvolutionWorks in more detail, focusing on the contributions listed in the introduction, and then providing additional details about the implementation.

The EvolutionWorks main window, as shown in Fig. 1, has a search dialog on the left and a *view* of the *workspace* on the right. The workspace is the entire contents of the visualization, while the current view shows only a portion of the workspace at one time. The background of the view is a grid that always displays about ten solid lines and about one-hundred fine lines that become more solid as the view zooms in. The fine lines eventually replace the original ten solid lines (the opposite happens when the view is zoomed out), which gives the impression of a grid with infinite gradation. The purpose of this infinite grid is to enhance the sense of motion as the view pans and
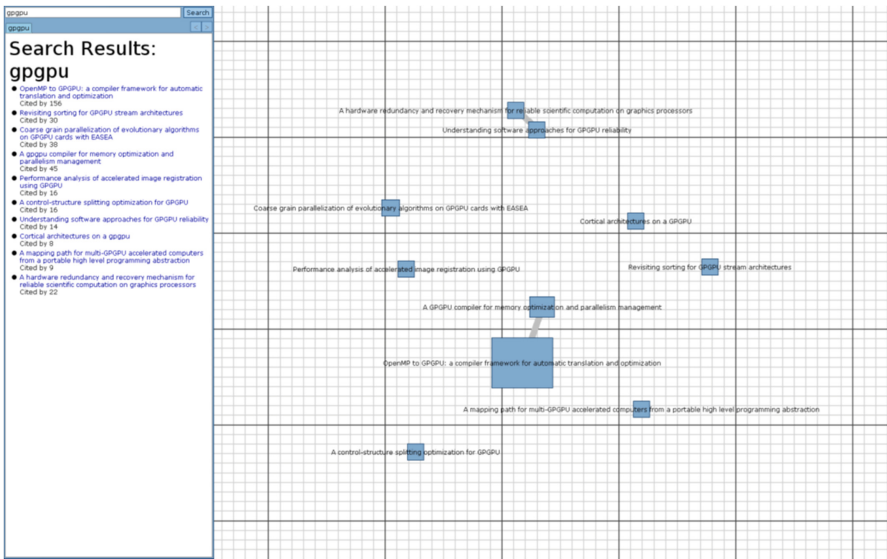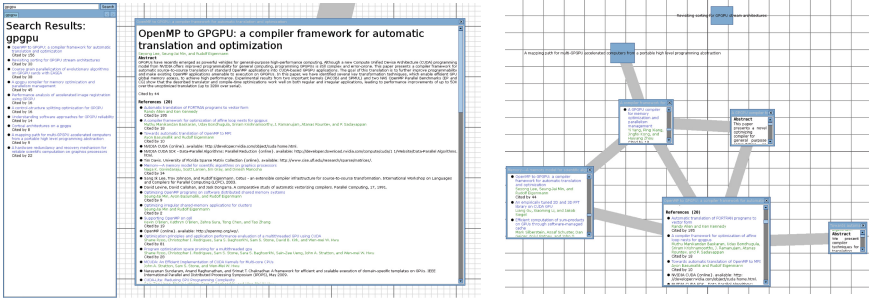

Fig. 1. The selected search results seed the workspace.

(a) When a paper gets the focus, it fills the view, but leaves a small margin.

(b) After a researcher selects references, they sprout out from the paper like leaves.

Fig. 2. Grow

zooms and additionally it provides the user feedback when there are no papers visible in the view.

Clicking on any paper link gives the target paper focus and causes the view of the workspace to smoothly pan and zoom so that the target paper becomes centered in the view and large enough to read as illustrated in Fig. 2a.

4.1 Seeding

The system initially presents the user with an empty workspace and a conventional text box for searching. Clicking on a search result opens a window in the workspace that contains the clicked paper's metadata. Entering the direct URL of a paper—say from the ACM Digital Library—is another way to open a window for that paper. The size of the paper's window is proportional to a selected metric, such as the number of citations, or the computed impact. If papers added to the workspace are related, then they are automatically linked by an edge (causing them to be pulled closer together by the force-directed layout), which is illustrated by two separate pairs of papers in Fig. 1.

To determine the placement of a paper in the workspace, the layout engine generates multiple random candidate locations within a bounding rectangle that surrounds all the other windows in the workspace. The layout engine weighs candidates according to the distance to other windows and to the edges of the view, and selects the best candidate. This method keeps windows from overlapping and keeps new windows close to existing ones as the number of windows grows.

4.2 Growth

While selecting a paper from the search dialog adds the paper to the workspace, selecting a new paper from a link in an existing paper causes the new paper to "sprout" from the existing paper like a leaf from a branch. As before, new edges connect the new paper to papers already in the workspace. The placement of a new citation or reference

window is determined as follows. A new window starts out overlapping with the existing window that it sprouts from with its center point just slightly offset from the center of the existing window. This initial placement determines what direction that the simulation will push the window after it is created. The repulsive force that keeps windows from overlapping will push the existing and new window away from one another until they are no longer overlapping, which gives the appearance that the new window grew out of the exiting one. To determine the initial placement, the layout engine generates several candidate directions and weighs them according to how much free space there is to grow in that direction, and then it chooses the most unused space.

In Fig. 2b, the user has selected three references from the original paper in Fig. 2a, and as a result, three new papers are displayed. Further, the user has followed the chain one step further and selected a reference from one of the new papers, causing a fourth new paper to appear. Notice that the new papers also reference to each other, as indicated by the new edges between them.

4.3 Pruning

Pruning a single window from the workspace is simply a matter of clicking the *close* button on that window. The window and all of its associated edges will disappear from the workspace.

Additionally, user can remove an entire branch as long as there is only one edge connecting it to the rest of the graph. Hovering the mouse cursor over an eligible edge reveals a *delete* button on that edge, as shown in Fig. 3a. Clicking the button will prune the entire branch as shown in Fig. 3b. If the edge is not eligible to be pruned, due to not being the sole bridge to a sub-graph, then a circle-backslash will overlay the *delete* button to indicate that pruning is disabled. The user can work around this by pruning windows instead of edges.
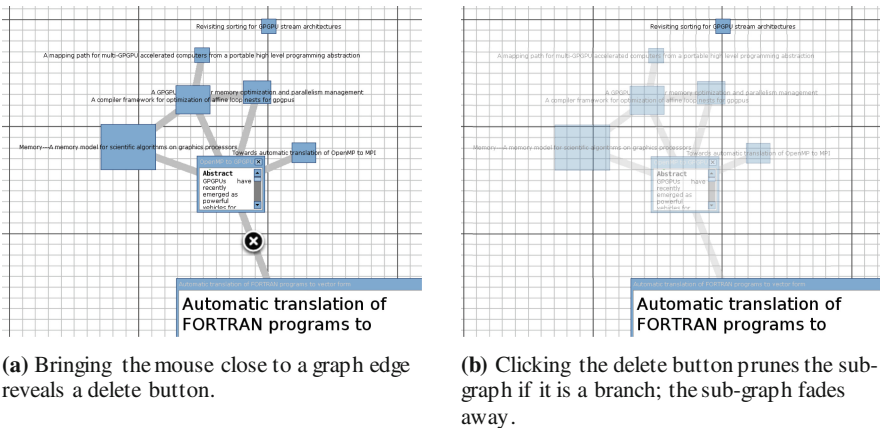
(a) Bringing the mouse close to a graph edge reveals a delete button.

(b) Clicking the delete button prunes the sub-graph if it is a branch; the sub-graph fades away.

Fig. 3. Prune

4.4   Kinetic Layering

With visual layering, retinal variables encode information, but with what we call *kinetic layering*, information is encoded in the physical properties of a physically based model. As retinal variables do with visual layering, in EvolutionWorks the motion of the elements in the simulation convey information *metaphorically*. We emphasize the term *metaphorical* here because physically based simulations are typically intended to be taken literally—but this is not the case for kinetic layering. For example, a literal simulation could mean, "this object weighs one thousand *kilograms*," however for kinetic layering the meaning of a physical property is metaphorical, so it might mean, "this object weights one thousand *citations*." As mentioned above, this is the same as with visual layering, which is also metaphorical; for example, in photography or representational art something red is literally red, but in a visualization, red is assigned a metaphorical meaning.

Some variables in a physically based model overlap with retinal variables, for example, location, size, shape, and orientation (particle systems only model location while rigid-body simulations model all of these). Typical physically based simulations use color, value, and texture either for pure aesthetics or metaphorically to convey information, while the physical variables are meant to be taken literally—a literal physical simulation with visual layering. The physical properties are things such as velocity, friction, rotation (angular velocity), and acceleration due to forces. Forces can be between objects and other objects, or forces can come from a force field (e.g. gravity). Kinetic layering uses these purely physical properties metaphorically, which leads to a physically based simulation where *nothing* displayed, neither retinal variables nor motion, is meant to be taken as a literal representation of physical reality.

EvolutionWorks uses kinetic layering by making papers with more citations larger and more massive—they push harder on other papers. Since the simulation is interactive, if a user attempts to pull an important paper around by pulling on a minor paper (like trying to lead a big dog on a leash), the important paper barely moves, which conveys its metaphorical weight. On the other hand, grabbing the important paper and moving it around easily flings about the referring papers because they are so light.

4.5   Focus-Context-Focus Hopping

To show as much metadata of a paper as possible, *EvolutionWorks* zooms in fairly close to a paper window that gets focus. This could make it difficult to keep the context of a paper in mind because all the other papers are outside of the view. This is why the view is never panned directly from one paper to another (which would give the impression of passing by a picket fence and looking through it closely), but instead the view is zoomed out first to reveal the context. For example, if the user clicks on a reference in the paper shown in Fig. 2a, the view zooms out to reveal the relationship between the original focus and the next focus. Then, as shown in Fig. 4a, the view pans over to center the selected paper, as shown in Fig. 4b, and then zooms in to focus on the selected paper, as shown in Fig. 4c.
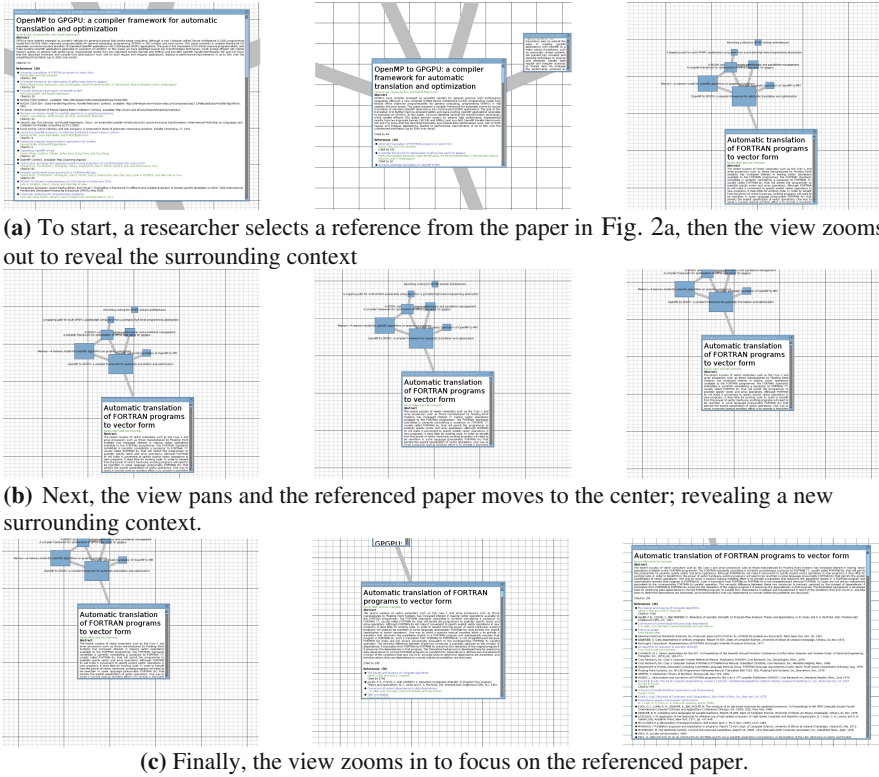
(a) To start, a researcher selects a reference from the paper in Fig. 2a, then the view zooms out to reveal the surrounding context

(b) Next, the view pans and the referenced paper moves to the center; revealing a new surrounding context.

(c) Finally, the view zooms in to focus on the referenced paper.

Fig. 4. Focus-context-focus hop

4.6   Unified Presentation

Except for the initial search text box, *EvolutionWorks* displays everything in a single unified view. The nodes of the graph are windows containing paper metadata. This presents a couple of challenges, the first being aesthetics, and the second being performance.

Full windows do not look pleasing when scaled down to a small size. Therefore we use a *semantic zoom* [34] that changes windows to a simpler representation depending on their size. Our current implementation only has two levels of semantic zoom—the full window and a simple rectangle with an overlaid title.

When many windows are open, the computational requirements for smooth animation can become overwhelming. Resizing dozens of full windows that contain text that needs to be word-wrapped at sixty frames per second is challenging even for powerful machines. *EvolutionWorks* thus uses a simplified window representation while the view is animating. This way the layout engine only needs to be fully reformat text once the view stops changing. The simplified presentation preserves information, such as the context, during a focus-context-focus hop, but it does not spend time re-computing information that users would most likely not be able to read during animation anyway.

4.7   Cluster Title Summarization

Figure 5a shows how difficult it can be to make out individual paper titles once the view is zoomed out far enough, as titles can overlap, and become illegible. EvolutionWorks creates a more concise representation for papers when they become jumbled like this. Small clusters of papers are detected by their proximity in the view. The titles of the papers in the same cluster are then replaced by a selection of common terms that appear in the metadata of cluster members. A set of common stop words like "the" are ignored. The *Porter stemming algorithm* [35] is used to factor out different forms of the same word. The *tf-idf* [36] score of the words appearing in each cluster are counted, and the top six terms are concatenated to make a new title for each cluster, like the ones shown in Fig. 5b.
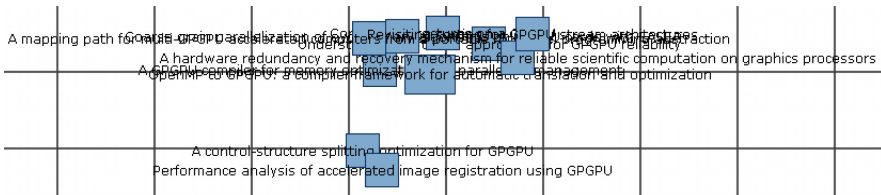
(a) Showing full titles of many papers at once in a small space can become an illegible mess.
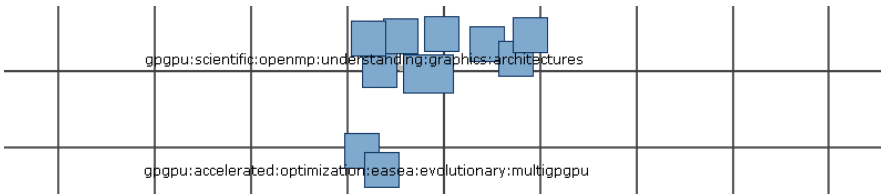
(b) Titles are replaced by a list of the most prevalent words, which keeps the view legible.

Fig. 5. Cluster title summarization

4.8   Implementation

EvolutionWorks relies on several technologies to enable its simulation, animation, and graphics. We use *Gephi* [37, 38] for manipulating graphs and for physically based modeling, and *OpenGL* [39] with the *Themable Widget Library* (TWL) [40] for the user interface. Gephi is a powerful graph visualization application and library for Java—described as "Photoshop for graphs." Specifically, we use Gephi's *ForceAtlas2* layout algorithm for small to medium sized graphs. Additionally, there is a plug-in system for defining new layout algorithms. For visuals, OpenGL and TWL provide the hardware accelerated high performance graphics required for kinetic layering and the unified presentation. Standard Java widget libraries have significant overhead, which is unacceptable for implementing smooth animation. TWL puts a lot of emphasis on efficiency because its purpose is for use in video games. Since TWL is implemented with OpenGL, combining TWL user interface elements with OpenGL graphics primitives is natural.

Another important capability that EvolutionWorks needs is structured access to the Web. For this, EvolutionWorks uses *BigSemantics* [41], an open source system that can provide structured semantic information from the Web even though the Web is only semi-structured. We use the object-oriented *Meta-Metadata* language, provided by BigSemantics, to describe *semantic data types* [42] for scholarly articles. While different digital libraries and repositories—such as Google Scholar [43], ACM Digital Library [44], IEEE Xplore [45], and CiteSeer [46]—use different HTML representations, BigSemantics provides a common base type that allows EvolutionWorks to use scholarly articles from different sources as if they were from the same source.

5 User Study

To evaluate whether *EvolutionWorks* improves the experiences of exploratory browsing of academic literature, we devised a user study that simulated a hypothetical scenario. We asked the participants to make a short reading list of five papers to recommend at an imaginary future lab meeting. This activity is an information-based ideation task [5] that requires users to think iteratively about the research questions and collect significant ideas. Our hypothesis was that users of EvolutionWorks would produce reading lists with a higher citation count than those made by users of a regular Web browser. We chose the citation count, as retrieved from the ACM Digital Library, as the metric to measure the quality of reading lists. Although the number of experimental trials was small, we found a marked increase in the number of citations attributed to the papers selected using EvolutionWorks. Citation counts correlate with identification of significant research.

5.1 Experimental Design

We created two separate tasks and two separate groups, which resulted in there being four subgroups. The first task was to look up papers relevant to human computer interfaces and *Healthcare*, and the second task was to look up papers relevant to human computer interfaces involving the *Wiimote* controller (by *Nintendo*). The subgroups used either a version of EvolutionWorks that looks similar to a conventional tabbed *Browser*, or the full version that uses EvolutionWorks' *Graph* visualization. The result is four subgroups, *Healthcare-Browser*, *Healthcare-Graph*, *Wiimote-Browser*, and *Wiimote-Graph*. To increase the amount of trials, we asked some users to participate in two subgroups with different tasks. For example, if a user had completed *Healthcare-Browser* then they could also complete *Wiimote-Graph*. To counter any learning effect we randomly determined the order that the users would undertake the *Browser* or *Wiimote* tasks.

5.2 Hypothesis

The size and weight given to papers with more citations should draw the user's attention to these papers more than in a list of papers sorted by citation count. The

difference should be noticeable in the number of citations attributed to the papers found by users that are in the Graph group. In addition, the mean time taken to complete a reading list by the Graph group should not be significantly different from the mean time of the Browser group.

5.3   Results

It turned out that the mean of the time taken by the users in each group was not statistically significant. For the *Wiimote* task, the browser subgroup happened to take less time; the mean time taken of the *Wiimote-Browser* subgroup was 5 min 24 s and that of the *Wiimote-Graph* subgroup was 10 min and 51 s. The opposite was observed of the *Healthcare* task with the *Healthcare-Graph* subgroup seeming to do better by taking a mean time of 7 min and 24 s and the *Healthcare-Browser* group a mean time of 9 min and 14 s. The two-tailed P value for these means was 0.3952, which indicates a 40 % probability that these results were due to variability in the measurements and are not significant. This is similar to the results of tests of the XML3D [33] browser where they found that the use of a non-conventional visualization did not cause a significantly affect the time spent on browsing. Users seem to work until they get bored with a task and want to move on, so any improvement in results is due to more efficient exploration, and not because the user spent more time with EvolutionWorks.

Figures 6 and 7 visualize the results of the user study as a histogram. Each bar represents a paper whose height is based on the citation count scaled by the number of users in each group. In Fig. 8 the same histograms are overlaid on top of one another to
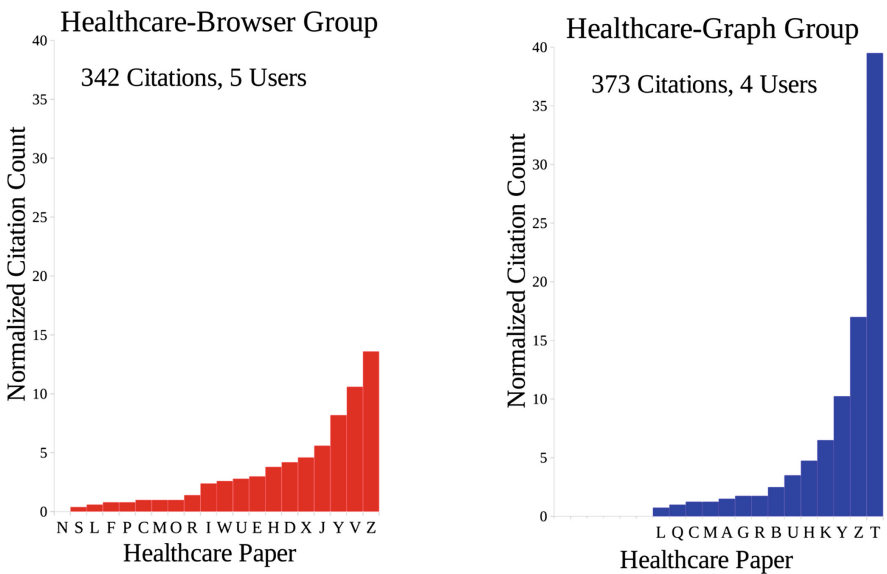

Fig. 6. The histograms for the Wiimote task indicate the large advantage we observed that the Graph group had over the Browser group in selecting papers that are more highly cited. The bars in each histogram labeled with the same letter or digit correspond to the same paper.
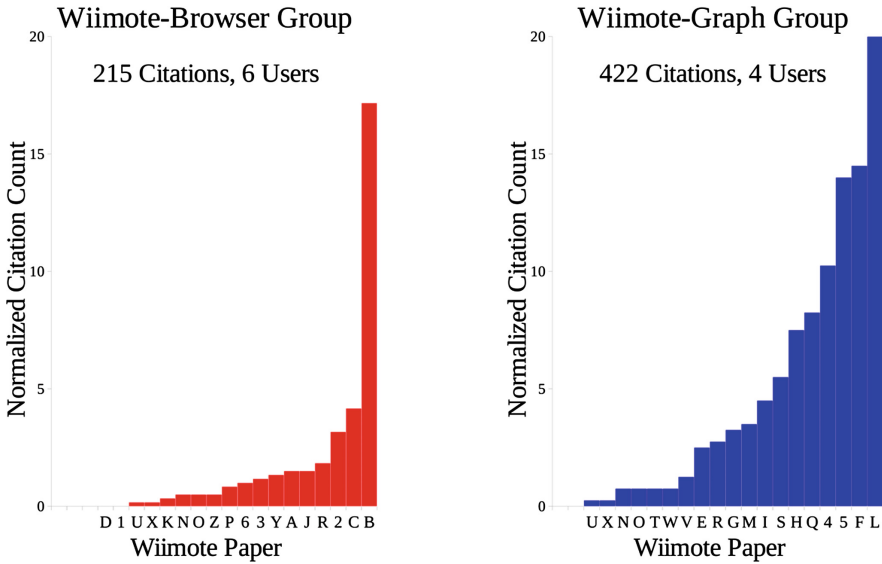
Fig. 7. The advantage of the *Graph* group is smaller for the *Healthcare* task than for the *Wiimote* task shown in Fig. 6, but it is still clearly observed. The bars in each histogram labeled with the same letter correspond to the same paper.
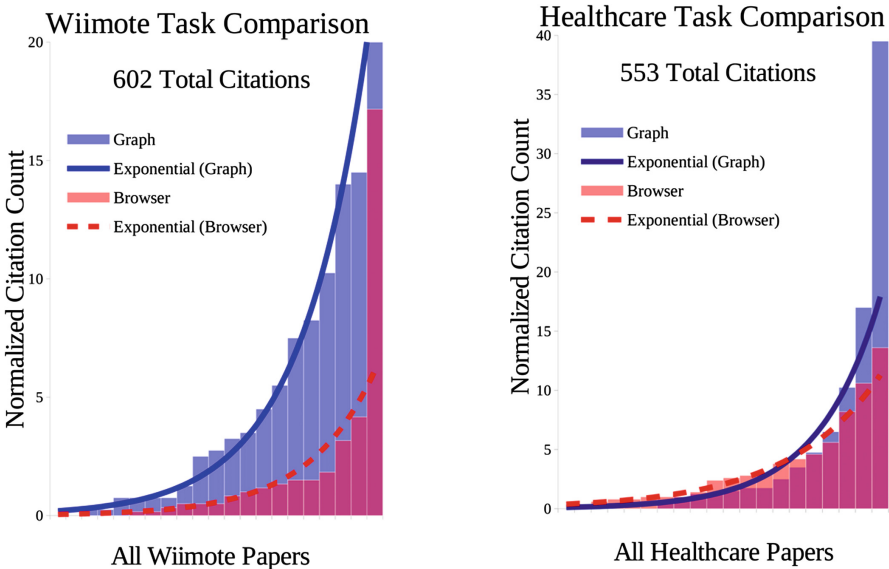

Fig. 8. Superposition of the histograms from Figs. 6 and 7. Exponential curves have been fitted to the data to show more clearly the difference in trends.

demonstrate the differences. To help show the difference in the results more clearly, we have fit exponential curves to the histograms. An exponential curve was chosen

because it has been shown empirically [47] that the distribution of citation counts in the academic citation network is consistent with a power law governing how quickly the number of citations of a paper grows over time. The general impression is that most of the papers found by the *Graph* group are of higher quality than those found by the *Browser* group, because the area under the curve is greater.

6   Conclusion and Future Work

After interviewing researchers about their methods of researchers conducting literature searches, we noticed a pattern that we dubbed *seed-grow-prune*. The *seed* is a research idea that suggests keywords for the researcher to search the web or an online public access catalog. From those initial results, the researcher follows link after link (prioritizing based mostly on paper titles alone), forming a chain, so that eventually the seed ends up *growing* into an unwieldy pile. After looking at the abstracts and contents of papers in more detail, the researcher *prunes* many of the papers away until there is finally a manageable reading list.

The novel features of EvolutionWorks, including: *kinetic layering*, a *unified presentation*, *focus-context-focus hopping*, and *cluster title summarization*, work together to support seed-grow-prune. First, *kinetic layering* allows growth to take place smoothly while conveying abstract document properties in the motion of a physically based simulation. The continuous plant-like motion is compelling, avoids change blindness, and gives a literal meaning to seed-grow-prune. Second, the *unified presentation* displays documents directly in the graph in order to avoid a split in the researcher's attention between the graph and a separate document viewer. Third, *focus-context-focus hopping* keeps the broader context of documents in the mind of the researcher when switching focus between papers. Without hopping, the user might pan from paper to paper in the view without ever seeing much context and lose the benefit of the unified presentation. Finally, *cluster title summarization* keeps the view tidy by simplifying the titles of papers that are too small (in the current view) for the researcher to examine individually, which can happen because the importance metric can cause the size of different papers to vary widely.

To evaluate EvolutionWorks we conducted a user study that gave promising results, but also provides seeds of ideas for future research. The study reported in this paper was small, but we obtained positive indicative results that the quality of papers found by users of EvolutionWorks is higher than that of the papers found with a traditional Web browser. However, for stronger confidence in the benefits of EvolutionWorks, in the future a larger user study is necessary—especially to pinpoint the value of each individual feature. In particular, the benefit of different physical properties used with kinetic layering needs to be further studied. So far, we have purposefully not taken advantage of most retinal variables, which allows us to avoid confusing the benefits of visual layering with the benefits of kinetic layering. We say most, because size is both a retinal variable and a physical property at the same time. Since size is actually the most dominant retinal variable, we need to measure the benefit of kinetic layering where there are no size differences. Another possible experiment could isolate even the effects of paper titles and citation counts by showing that kinetic

layering alone is enough to influence users to select papers based on a "quality" metric that assigns papers completely random values.

Another avenue of future work is the application of filtering to kinetic layering. A visualization system normally displays filtered results by making marked variations in visual layering, such as changing the size, color, brightness, etc., of objects. Extending this concept to kinetic layering would give us *kinetic filtering*, where the behavior of a simulation could be modified dynamically by changing the physical properties of objects based on a filter.

Finally, we would like to apply EvolutionWorks to other forms of information-based ideation other than using the academic citation network to build reading lists. *BigSemantics* already includes formats for many other kinds of creative works and social media. Consequently, the system could bring multiple kinds of data into the view at once, such as *Wikipedia* articles, news stories, tweets, social media profiles, etc. — all grown from a user's seed — all connected together in a single workspace.

References

1. Herman, I., Melançon, G., Marshall, M.S.: Graph visualization and navigation in information visualization: a survey. IEEE Trans. Visual Comput. Graphics **6**, 24–43 (2000)
2. Heer, J., Boyd, D.: Vizster: visualizing online social networks. In: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization, pp. 532–539. IEEE Computer Society, Washington, DC, USA (2005)
3. Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J.J., Fekete, J.-D., Fellner, D.W.: Visual analysis of large graphs: state-of-the-art and future research challenges. Comput. Graphics Forum **30**, 1719–1749 (2011)
4. Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, Madison (1983)
5. Kerne, A., Webb, A.M., Smith, S.M., Linder, R., Lupfer, N., Qu, Y., Moeller, J., Damaraju, S.: Using metrics of curation to evaluate information-based ideation. ACM ToCHI **21**, 48 pp. (2014)
6. Jain, A., Lupfer, N., Qu, Y., Linder, R., Kerne, A., Smith, S.M.: Evaluating TweetBubble with ideation metrics of exploratory browsing. In: Proceedings of Creativity and Cognition. ACM (2015)
7. Mostafa, J.: Seeking better web searches. Sci. Am. **292**, 66–73 (2005)
8. Fast, K.V., Campbell, D.G.: "I still like Google": University student perceptions of searching OPACs and the web. Proc. Am. Soc. Inf. Sci. Technol. **41**, 138–146 (2004)
9. Griffiths, J.R., Brophy, P.: Student searching behavior and the web: use of academic resources and Google. Libr. Trends **53**, 539–554 (2005)
10. Bell, S.J.: The Infodiet: how libraries can offer an appetizing alternative to Google. Chronicle Higher Educ. **50**, B15 (2004)
11. Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H.R., Dobrowolski, T., Tenopir, C.: The Google generation: the information behaviour of the researcher of the future. Aslib Proc. **60**, 290–310 (2008)
12. Bates, M.J.: The design of browsing and Berrypicking techniques for the online search interface. Online Rev. **13**, 407–424 (1989)

13. Koh, E., Dworaczyk, B., Albea, J., Hill, R., Choi, H., Caruso, D., Graeber, R., Mistrot, J.M., Smith, S.M., Webb, A., Kerne, A.: combinFormation: a mixed-initiative system for representing collections as compositions of image and text surrogates. In: Joint Conference on Digital Libraries, vol. 0, pp. 11–20 (2006)

14. Eades, P.: A heuristics for graph drawing. Congressus Numerantium **42**, 146–160 (1984)

15. Fruchterman, T.M., Reingold, E.M.: Graph drawing by force-directed placement. Softw. Pract. Experience **21**, 1129–1164 (1991)

16. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. Inf. Process. Lett. **31**, 7–15 (1989)

17. Frick, A., Ludwig, A., Mehldau, H.: A fast adaptive layout algorithm for undirected graphs (extended abstract and system demonstration). In: Tamassia, R., Tollis, I.G. (eds.) GD 1994. LNCS, vol. 894, pp. 388–403. Springer, Heidelberg (1995)

18. Tufte, E.R.: Envisioning information. Optometry Vision Sci. **68**, 322–324 (1991)

19. Lum, E.B., Stompel, A., Ma, K.-L.: Kinetic visualization: a technique for illustrating 3D shape and structure. In: Visualization, pp. 435–442 (2002)

20. Wallach, H., O'Connell, D.N.: The kinetic depth effect. J. Exp. Psychol. **45**, 205–217 (1953)

21. Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **14**, 201–211 (1973)

22. Simons, D.J., Levin, D.T.: Change blindness. Trends Cogn. Sci. **1**, 261–267 (1997)

23. Elmqvist, N., Tsigas, P.: CiteWiz: a tool for the visualization of scientific citation networks. Inf. Visual. **6**, 215–232 (2007)

24. Elmqvist, N., Tsigas, P.: Causality visualization using animated growing polygons. In: IEEE Symposium on Information Visualization, pp. 189–196 (2003)

25. Bergstrom, P., Atkinson, D.C.: Augmenting the exploration of digital libraries with web-based visualizations. In: Fourth International Conference on Digital Information Management, pp. 1–7 (2009)

26. Keim, D.A., Schneidewind, J.O., Sips, M.: CircleView: a new approach for visualizing time-related multidimensional data sets. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 179–182. ACM, Gallipoli, Italy (2004)

27. Small, H.: Co-citation in the scientific literature: a new measure of the relationship between two documents. J. Am. Soc. Inf. Sci. **24**, 265–269 (1973)

28. Chen, T.T., Hsieh, L.C.: On visualization of cocitation networks. In: Proceedings of the 11th International Conference on Information Visualization pp. 470–475. IEEE Computer Society (2007)

29. Chen, T.T., Yen, D.C.: CociteSeer: a system to visualize large cocitation networks. Electron. Libr. **28**, 477–491 (2010)

30. Blythe, J., Patwardhan, M., Oates, T., desJardins, M., Rheingans, P.: Visualization support for fusing relational, spatio-temporal data: building career histories. In: Proceedings of the 9th International Conference on Information Fusion, pp. 1–7 (2006)

31. Morris, S.A., Yen, G., Wu, Z., Asnake, B.: Time line visualization of research fronts. J. Am. Soc. Inform. Sci. Technol. **54**, 413–422 (2003)

32. Chou, J.-K., Yang, C.-K.: PaperVis: literature review made easy. Comput. Graphics Forum **30**, 721–730 (2011)

33. Risden, K., Czerwinski, M.P., Munzner, T., Cook, D.B.: An initial examination of ease of use for 2D and 3D information visualizations of web content. Int. J. Hum. Comput. Stud. **53**, 695–714 (2000)

34. Perlin, K., Fox, D.: Pad: an alternative approach to the computer interface. In: Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, pp. 57–64. ACM, Anaheim, CA (1993)

35. Van Rijsbergen, C.J., Robertson, S.E., Porter, M.F.: New models in probabilistic information retrieval. Computer Laboratory, University of Cambridge (1980)
36. Jones, K.S.: A statistical interpretation of term specificity and its application in retrieval. In: Peter, W. (ed.) Document Retrieval Systems, pp. 132–142. Taylor Graham Publishing, London (1988)
37. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks. In: Third International AAAI Conference on Weblogs and Social Media, pp. 361–362. AAAI Publications (2009)
38. Gephi. http://gephi.github.io. Accessed 12 May 2015
39. OpenGL. http://opengl.org. Accessed 12 May 2015
40. TWL—Themable Widget Library. http://twl.l33tlabs.org/. Accessed 12 May 2015
41. Kerne, A., Qu, Y., Webb, A.M., Damaraju, S., Lupfer, N., Mathur, A.: Meta-metadata: a metadata semantics language for collection representation applications. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1129–1138. ACM, Toronto, ON, Canada (2010)
42. Qu, Y., Kerne, A., Lupfer, N., Linder, R., Jain, A.: Metadata type system: integrate presentation, data models and extraction to enable exploratory browsing interfaces. In: Proceedings of EICS. ACM (2014)
43. Google Scholar. http://scholar.google.com. Accessed 12 May 2015
44. ACM Digital Library. http://dl.acm.org. Accessed 12 May 2015
45. IEEE Xplore. http://ieeexplore.ieee.org. Accessed 12 May 2015
46. CiteSeer. http://citeseer.ist.psu.edu/index. Accessed 12 May 2015
47. Redner, S.: How popular is your paper? An empirical study of the citation distribution. Eur. Phys. J. B Condens. Matter Complex Syst. **4**, 131–134 (1998)