Visual Interfaces for Recommendation Systems: Finding Similar and Dissimilar Peers

FAN DU, CATHERINE PLAISANT, NEIL SPRING, and BEN SHNEIDERMAN, University of Maryland, USA

Recommendation applications can guide users in making important life choices by referring to the activities of similar peers. For example, students making academic plans may learn from the data of similar students, while patients and their physicians may explore data from similar patients to select the best treatment. Selecting an appropriate peer group has a strong impact on the value of the guidance that can result from analyzing the peer group data. In this paper, we describe a visual interface that helps users review the similarity and differences between a seed record and a group of similar records, and refine the selection. We introduce the LikeMeDonuts, Ranking Glyph, and History Heatmap visualizations. The interface was refined through three rounds of formative usability evaluation with 12 target users and its usefulness was evaluated by a case study with a student review manager using real student data. We describe three analytic workflows observed during use and summarize how users' input shaped the final design.

CCS Concepts: • **Human-centered computing** → **Visual analytics**; *Graphical user interfaces*; • **Information systems** → **Recommender systems**; *Similarity measures*;

Additional Key Words and Phrases: Similarity, personal record, multidimensional data visualization, temporal visualization, decision making, visual analytics

ACM Reference Format:
Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2018. Visual Interfaces for Recommendation Systems: Finding Similar and Dissimilar Peers. *ACM Trans. Intell. Syst. Technol.* 1, 1, Article 1 (January 2018), 23 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Recommendation applications can guide users in making important life choices by referring to the activities of similar peers. With the rapid accumulation and digitization of personal records, software tools have been developed to enable the retrieval and analysis of the data of similar individuals to facilitate making important decisions. For example, patients and their physicians may explore data from similar patients to select the best treatment (e.g., PatientsLikeMe [51], CureTogether.com). Students making academic plans may be inspired by the achievements of similar students (e.g., PeerFinder [13], EventAction [12]). While automated black-box recommendation techniques are effective and used widely in shopping and entertainment applications [19, 30, 41], transparency is critical when users review data and recommendations for life decisions, carefully decide to accept a recommendation, or remain doubtful [23, 44]. In this paper, we focus on how to improve the selection of peer groups, i.e., how to select "people like me," or "people like the patient, student, or customer I am advising."

Authors' address: Fan Du; Catherine Plaisant; Neil Spring; Ben Shneiderman, University of Maryland, College Park, MD, 20742, USA, {fan,plaisant,nspring,ben}@cs.umd.edu.

2157-6904/2018/1-ART1 $15.00
https://doi.org/0000001.0000001

Previous work suggested that users are more engaged and more confident about making important life choices when provided with more controls and more context, even at the cost of increased complexity [13]. The next question then becomes: which controls and context should be provided? How do users find a satisfying peer group? And how can we facilitate this process?

In this paper, we report on three visualization designs and three analytic workflows to support users in retrieving, reviewing, and refining peer groups, making use of both record attributes and simple patterns of temporal events found in the record. We introduce LikeMeDonuts, a novel hierarchical visualization providing an aggregated overview while preserving details about individual peers, to support users in reviewing similarities and differences of a group of records compared to the seed record. While most existing tools focus on hierarchies that have a fixed structure (e.g., the ICD-10 codes [55] or phenotypes [18]), we investigate situations when the order of the hierarchy is flexible and subjective, depending on the analysis goals and users' preferences. Our prototype provides controls for users to interactively adjust the layout, create visual representations that best satisfy their needs, and refine the peer group composition. It also provides recommendations on improving the layout so as to reduce visual clutter and mitigate issues of scalability.

We refined the design through three rounds of formative usability evaluation with a total of 12 target users, and report how the prototype evolved on users' feedback. We propose three analytic workflows for forming peer groups and report on users' experience and preferences.

Our contributions include:

- A novel hierarchical visualization (LikeMeDonuts) that provides an overview of peer groups with a flexible hierarchy of criteria values, similarity encoding, and interactive support for trimming the peer group.
- An interactive visualization system (iteratively refined through three rounds of formative usability study) that combines three new visualization components and supports three analytic workflows.
- A case study conducted with a graduate student review manager using real student data to evaluate the usefulness of the design.

2 BACKGROUND AND RELATED WORK

Modelling personal records. Personal records are typically modeled as having record attributes such as gender, age, and education level. They are usually complemented by temporal event data representing the persons' activities in a period or milestones that occurred in their life [38]. Connections between individuals [40, 56] can also be analyzed but will not be addressed in this paper. Similarity is a fundamentally important concept in many research domains [1]. For example, in bioinformatics for gene sequence alignment [28] or protein clustering [29], in linguistics for approximate string matching [35] or text categorization [4], in computer vision for face recognition [42], and in healthcare for identifying similar patients [48, 51]. Some approaches may use a single criterion, such as the presence of a diagnosis [51] but most approaches use black-box models to assess similarity and perform the search in an automated way, while we provide user control over the process and facilitate the interactive review and refinement of the results.

Closest related work. The closest related paper describes PeerFinder [13], which was designed for finding similar people using both record attributes and temporal history. The paper discussed the challenges of finding similar people, described the initial interface and provided evidence for the benefit of increasing user control and context information, but it did not propose any novel visual designs to fulfill this need. Our paper builds on this early work, documents the evolution of the designs, and reports on their use.

The introduced interface design falls into the general category of multidimensional data visualization since it handles all attributes of the records in a given dataset. Several techniques are commonly used in visualization tools to show record attributes in personal records. The icicle plots and sunburst layouts—its equivalent in polar coordinates, are used to show the distributions of attribute values in a group (e.g., PhenoStacks [17], PhenoBlocks [18], and InfoZoom [47]). Treemaps [2] are useful to showing the composition of a group based on (only) two attributes mapped to size and color (e.g., COQUITO [27]), but other attributes can be used to interactively compose the hierarchy [8]. Parallel coordinates can be used, or their cousin the radar plots [10]. Using tightly coupled sets of barcharts has also been explored [58].

Hierarchy. Hierarchies are powerful tools to represent complex data [14]. While most applications model record attributes as tree hierarchies that have a fixed structure (e.g., the ICD-10 codes [55] or phenotypes [18]), the exploration of similarity criteria benefits from the ability to customize the hierarchy. Compared to a classic sunburst, our LikeMeDonuts structure has three unique features. First, LikeMeDonuts are built on a tree of fixed depth with a reorder-able hierarchy of independent attributes, each of which makes a mutually exclusive covering over the set of items (e.g., "Program" splits graduate students into M.S. or Ph.D.). Second, LikeMeDonuts provide a set of operations for managing the display. Users can add and remove criteria from all interface components, and dynamically reorder the hierarchy based on specific analyses and their preferences, for example, allowing users to move a "Body Weight" attribute to the first level when looking for similar diabetic patients. Third, the photo at the center provides a visual reminder that all the information is relative to that person. The thickness of each donut ring and the color of each cell are meaningful in achieving the goal of finding similarity or differences.

Temporal patterns. To handle personal records, the interface also must incorporate temporal criteria: in our design, temporal patterns of interest are searched and new attributes are added to the records, reflecting whether the pattern has been matched or not, or if only a similar pattern was found. The new attributes and the quantitative similarity measures are added to the set of criteria, the LikeMeDonuts, and all the other visual components. We believe this unique combination of record attributes and temporal patterns enables users to review and understand the results in terms of personal attributes and personal history data, and compare to the seed record in a new way.

Search. Finding similar people can be seen as a straightforward search task: looking for records that exactly match a set of query rules. Standard query languages (e.g., TQuel [46] and T-SPARQL [20]) and interactive visual tools (e.g., (s|qu)eries [59], COQUITO [27], and EventFlow [34]) can be used to perform this task. However, we know that no two people are identical when using a rich set of attributes (with the possible exception of identical twins at birth), so the result set of identical people is typically null in datasets of interest. Users have to set a range of acceptable values for each attribute, i.e., the search criterion. This task has been well tackled by dynamic queries [25, 43] and faceted search interfaces [49, 52]. What we believe is a unique contribution in this paper is that we help users specify a small number of levels of similarity during the search process and present the results accordingly. Our prototype visually classifies records' attribute values as either (1) identical to the value of the seed record (bright green), (2) within acceptable range (dark green), (3) out of range (gray), or (4) excluded (i.e., filtered out so that records with those values are not visible). This general classification simplifies the results, but users need to see the actual value of all attributes to judge the results and are provided with ways to specify and adjust this grouping in the four categories. The LikeMeDonuts summarizes results of the search using a flexible hierarchy of criteria values, combined with a strong visual mapping of similarity and differences.

Ranking. Another characteristic of searching for similar records is that records in the result set are ranked, and users can decide where the cut-off is or how many records to keep. We introduce a

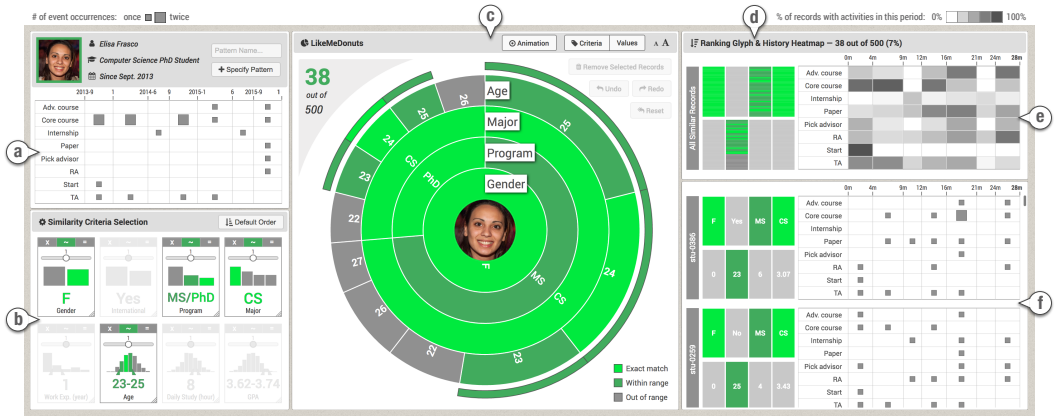

Fig. 1. The interface of our prototype for forming peer groups: (a) seed record timeline, (b) similarity criteria controls, (c) LikeMeDonuts representing criteria values of the 38 most similar records as a hierarchical tree, (d) Ranking Glyph providing a compact overview of 38 most similar records ranked by similarity, (e) History Heatmap showing the popularity of the temporal events among similar records, and (f) ranked list of similar records, displaying detailed information of individual records.

Ranking Glyph to address this need. Its design is inspired by Value Bars [7, 21] and pixel-based visualization in general [11, 26, 45].

People, not objects. Finally, a characteristic which is specific to our interface is that it deals with people, as opposed to objects like books, cars, or shoes. Prior work suggests that people (e.g., students and patients) express strong opinions when judging whether personal records are similar or dissimilar to them, influenced by their experience or beliefs [13], and has been described as a "slippery notion" [9]. This powerful subjective and personal component in determining similarity obliges designers to provide adequate control and context, thereby encouraging user engagement and inspiring trust in the results [13].

Comparing. Comparing just two records can be straightforward, using juxtaposition, superposition, or explicit encoding [16]. Special designs are available for complex cases such temporal event sequences [54], pairs of similar medication lists [39], or entire patient phenotypes [18]). Showing differences between even just a handful of records or patterns becomes more difficult [5, 15, 31, 61] but revealing the range of similarity and differences among a larger group of records presents substantial challenges that have rarely been studied (e.g., CoCo [32] for event sequences).

In summary, the carefully coordinated set of visual techniques proposed in this paper (LikeMeDonuts, Ranking Glyph, and History Heatmap) and the use of a consistent encoding of similarity enables users to interactively evaluate the similarities and differences of a ranked set of similar records compared to a seed record.

3 DESCRIPTION OF THE USER INTERFACE

After describing motivation and goals, this section describes the final design of the interface. The rest of the paper will describe early designs, problems uncovered during three rounds of usability testing, and how the design evolved. Finally, the discussion section addresses remaining challenges and possible solutions.

3.1 Motivations and Needs Analysis

In PeerFinder [13], we described user studies investigating how the complexity of the interface affects users' engagement in the decision making process and confidence in the results. We used two visualization components, barcharts and a ranked list, and evaluated the interface through a user study with 18 university students and interviews with 4 domain experts (three student advisors and a physician). Based on our discussions with the participants, we identified two critical users' needs which motivate the design of the new interface components introduced in this work:

N1. Tracking across multiple criteria. The interface should allow users to track and review a group of records that share similar values across multiple criteria, so that users can estimate the size of the group, explore how those records are distributed in other criteria, and refine the results by removing the group when necessary. The barcharts in our original design only support showing the value distribution of each separate criterion.

N2. Reviewing results at different levels-of-detail. The interface should provide both individual-level details and group-level overviews so that users can efficiently review and refine the results of similar records using both record attributes and temporal events. While the ranked list in our original design was useful to display full details of individual records, users were unable to get an overview of those records.

To satisfy these needs, we designed three new visualization components for reviewing and refining peer groups. Our main goal when designing LikeMeDonuts (Figure 1c) was to reveal distributions across combinations of multiple similarity criteria (e.g., female students majoring in computer science and having GPAs higher than 3.5). LikeMeDonuts allows users to estimate the size of multiple groups of records (i.e., the branches in a hierarchy of criteria) and provides interactive controls for selecting or removing groups, and rearranging the hierarchy that shapes those groups (**N1**).

The purpose of Ranking Glyph (Figure 1d) and History Heatmap (Figure 1e) was to provide a compact overview of the ranked list of the similar records (**N2**). The Ranking Glyph aimed to help users understand how similarities and differences for each criterion evolve as they go down the ranked list of similar records (e.g., are students having two internships more likely to be ranked on the top?). The History Heatmap helps users inspect common temporal patterns of activities for the entire peer group—or a selected subset (e.g., are students like me still taking classes in the fourth year?).

Those new components are integrated into the existing PeerFinder interface, which provides basic interface components: the seed record timeline (Figure 1a), similarity criteria controls (Figure 1b) and the underlying similarity search algorithm, and the basic ranked list of similar records for displaying detailed information (Figure 1f). Those basic components have also been refined as a beneficial side effect of the usability study (e.g., consistent use of color and improved coordination between components).

In the rest of this section, we describe the basic interface components first, then we present the new components in greater detail.

3.2 Basic Interface Components

Seed record timeline. The seed record's history of activities is shown as an aggregated timeline in a timetable (Figure 2a), where each row represents an event category and each column represents a time period. Events in each table cell are aggregated and represented as a square in gray and the number of event occurrences is represented by the size of the square. Users can specify temporal patterns of the seed record on the timeline and use them as similarity criteria for the search. In Figure 2, two temporal patterns have been specified based on the seed record's internship (having

Fig. 2. Four of the basic components that refine the PeerFinder interface: (a) seed record timeline, (b) similarity criteria controls, (c) similarity distribution, and (d) similar record distribution. In this example, a total of 10 similarity criteria are used, including two temporal criteria in the bottom row. The mouse cursor is hovering on the temporal criterion of "no papers in the first two years and late selection of an advisor." This criterion and the corresponding temporal pattern are highlighted in orange.

an internship every summer) and research activities (no papers in the first two years and late selection of an advisor). The temporal criteria are added as glyphs in the criteria control panel. Users can hover on a glyph to highlight the temporal pattern and the focused criterion in other visualizations in an orange color.

Similarity criteria controls. All available criteria are shown. Categorical criteria (such as major) and numerical criteria (such as GPA) are automatically extracted from the available data, and temporal criteria are added when specified by users. Each criterion is displayed as a rectangular glyph (Figure 2b) showing its name, the value for the seed record and the distribution for all archived records. Users can select how the criterion is to be used: "Ignore" (×), allow "Close Match" (~), or require "Exact Match" (=). A tolerance range can also be defined to treat multiple categorical

values or a range of numerical values as equivalent of the value of the seed record (e.g., treat M.S. and Ph.D. equally or set a GPA range between 3.2 and 3.7). The weight of each criterion can also be adjusted. As users adjust the controls, the results are updated immediately and reflected in all visualizations. Users can reorder the criteria by dragging the glyphs. Changes in order are reflected in other interface components but do not affect which records are included in the result set.

Similarity distribution. Based on the criteria settings, a similarity score is computed for each archived record (see PeerFinder [13] for algorithmic details) and a histogram of the scores is displayed (Figure 2c). Users can adjust the portion of the histogram that is selected for the results, i.e., the peer group. In Figure 2c, the top 20% most similar records (100 out of 500) are selected. Since the similarity scores change when users adjust the criteria controls, we provide three options to help users keep track of the record selection (shown as radio buttons in the toolbar): the "by Top N" option keeps users' selection of a fixed number of most similar records, the "by Percentage" option keeps the selection of a fixed percentage of most similar records, and the "by Similarity" option selects records whose similarity scores are above a user-defined threshold.

Similar record distribution. A separate view shows barchart distributions of criteria values of (only) the similar records (Figure 2d). The layout of the barcharts is consistent with the layout of the glyphs of the criteria control panel and the color of the bars is consistent with other components of the interface. Users can hover on a single bar to review the criterion range of values and number of records, and hover on a bar chart to highlight that criterion in other visualizations.

Basic ranked list of similar records. The individual records are displayed in a ranked list, showing the attribute values and the event history for each record (Figure 1f). For privacy, the individual records will need to be hidden when users do not have proper permission [13]. Part of the overviews or the labels may also need to be hidden when the number of records included is too low.

Improvements have been made to the basic interface components, e.g., the new color scheme used in the LikeMeDonuts was propagated to older components, and brushing and linking capabilities were added to coordinate all the views.

We now describe the new visualization components.

3.3 LikeMeDonuts

LikeMeDonuts is a radial space-filling visualization that shows the criteria values of the similar records as a hierarchical tree (Figure 3). An image of the seed record is placed at the center, anchoring the display on that person. Each donut ring represents a criterion (and one level of a tree structure). Criteria set to "Ignore" in the similarity criteria controls are not displayed. Ring sectors in bright green represent the proportion of people in the group whose values exactly match the value of the seed record, sectors in dark green represent those within the user-specified tolerance ranges, and gray sectors represent those outside tolerance ranges.

A thin additional partial ring is shown outside the donuts to highlight the records that are most similar to the seed record (based on the selected criteria). The arc is in bright green if the record's criteria values are all exactly matched, or in dark green if all criteria values are within range. When integrated into the larger interface, in Figure 4, we use the empty corner space to display contextual information and controls. The top left shows the number of similar records being reviewed and the total number of archived records. The color legend is at the bottom right. Controls for interactively editing the peer group within the LikeMeDonuts are at the top right corner.

3.3.1 Interactions. The donut rings and ring sectors are responsive to users' interactions and are linked to other visualizations on the interface. Hovering on a criterion in the similarity criteria controls highlights the matching donut ring with an orange border. Hovering on a ring sector highlights records represented by that sector with orange borders. When users click on one or

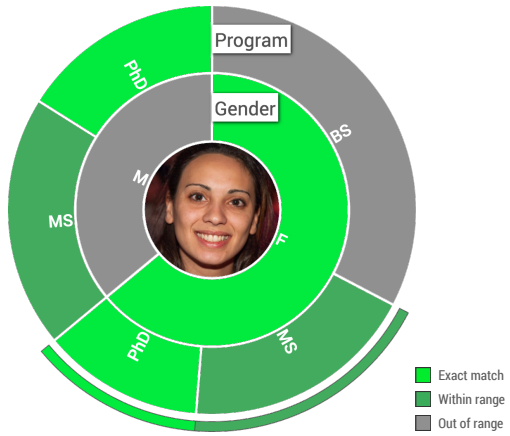

Fig. 3. This LikeMeDonuts shows two criteria as a two-level hierarchical tree. An image of the seed record is placed at the center. The inner ring represents gender. It shows that most records in the peer group are females like the seed record. The males are shown in gray, indicating that they are outside the tolerance range. The outer ring is for program. Among the females, most are B.S. students, and some are M.S. (shown in dark green because they are within range but not exactly like the seed record) or Ph.D. students. The males are all M.S. or Ph.D. students. The thin partial ring outside the donuts highlights records that are within range for both criteria.

multiple ring sectors, the selected records are highlighted in other visualizations (Figure 4): (1) orange bars are added in the similar record distribution barcharts, (2) the ranking of the selected records is shown in orange in the Ranking Glyph, (3) the History Heatmap shows the temporal activities of the selected records—using a color gradient from dark orange to white, (4) the individual selected records are be moved to the top of the ranked list of records with their IDs colored in orange, and (5) if a temporal criterion is used, the patterns will be highlighted with orange borders in the timelines of the similar records.

A set of control buttons are provided for editing the peer group at the record level. At the start, the buttons are disabled. Clicking on ring sectors will select a record subset and enable the "Remove Selected Records" button. As users make edits, the "Undo", "Redo", and "Reset" buttons become available. The removed records are filtered out and excluded in other visualizations immediately.

3.3.2 Animated Transitions. We carefully designed a four-stage animation [6, 22] to clarify the transition that occurs when users adjust the criteria controls or edit the peer group at the record level. The first stage fades out records removed from the peer group and criteria set to "Ignore" (i.e., removed). In the second stage, the LikeMeDonuts is resized to fill the screen space made available by removed donut rings or make space for new donut rings that will need to be added later. The third stage adjusts the size and color of the ring sectors and reorders them according to the updated peer group. The last stage fades in those newly added records and criteria/rings. A stage will be skipped if no changes occur during it. Each stage is set to 500 milliseconds. The entire animation takes two seconds at most for adjusting criteria controls, and one second for making an edit at record level (only involving the first and third stages). Users can turn the animation on or off.

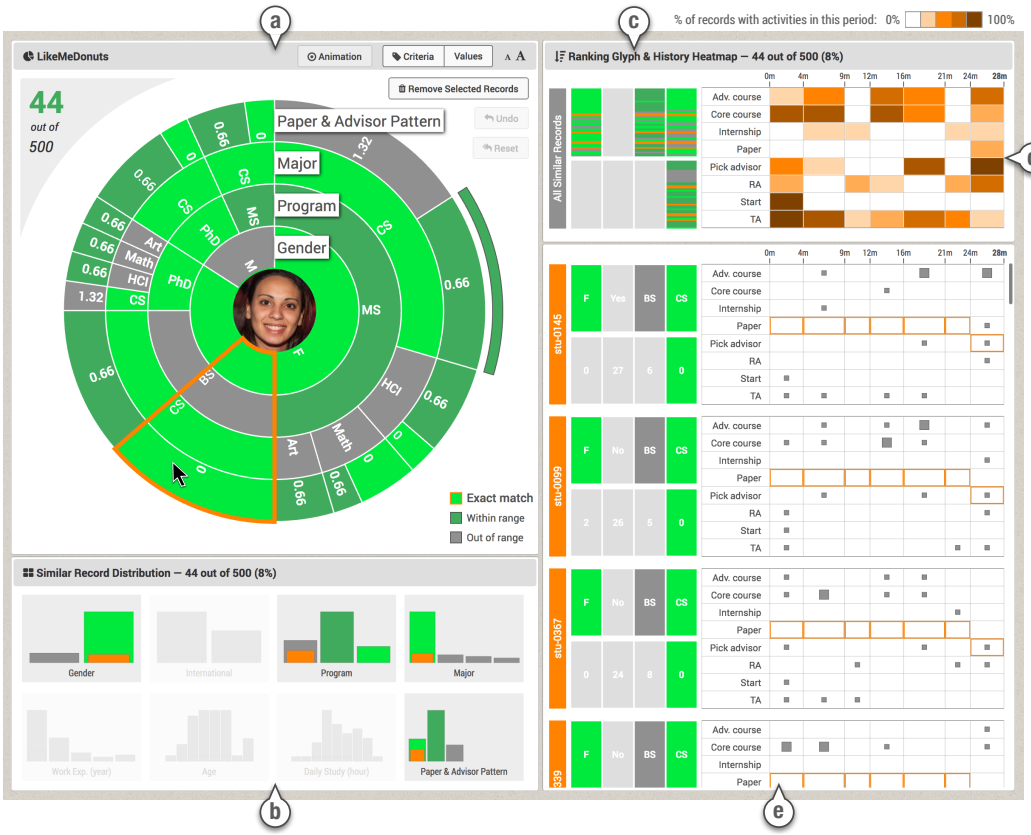

Fig. 4. All views are coordinated. In this example, a group of records are selected in the LikeMeDonuts (a) and therefore highlighted in orange in the similar record distribution (b), the Ranking Glyph (c) and the selected records are brought to the top of the similar record ranked list (e). The History Heatmap (d) is also updated to show only the events from the selected records. A "Paper and Advisor" temporal pattern was included in the criteria and appears as a numerical distance score in the LikeMeDonuts (with smaller values indicate more similar). The location of the pattern is also highlighted in the timelines of the individual records.

3.3.3 Order of Donut Rings. Given a set $\mathbf{C}$ of n criteria, the number of donut ring sectors is:

$$number\ of\ sectors = \sum_{i=1}^{n} \left(\prod_{j=1}^{i} \|c_j\| \right) \qquad c \in \mathbf{C}$$

where $\|c\|$ is the number of unique values of a criterion and as j increases, c_j moves from an inner ring to an outer ring. Note that $\|c_j\|$ appears in $(n - j + 1)$ terms of the summation. Therefore, inner rings have a larger impact on the result than outer rings. To minimize the number of sectors, criteria with smaller numbers of possible values should stay in the inner rings, whereas those with larger numbers of possible values need to be placed in the outer rings. Our system recommends an order of the donut rings at the start that minimizes the total number of sectors (therefore setting the default order of criteria in all other views). Users can then rearrange the rings to create views that better match their preferences by dragging the rings inward or outward, or dragging the criteria glyphs in the criteria control panel (Figure 1b).

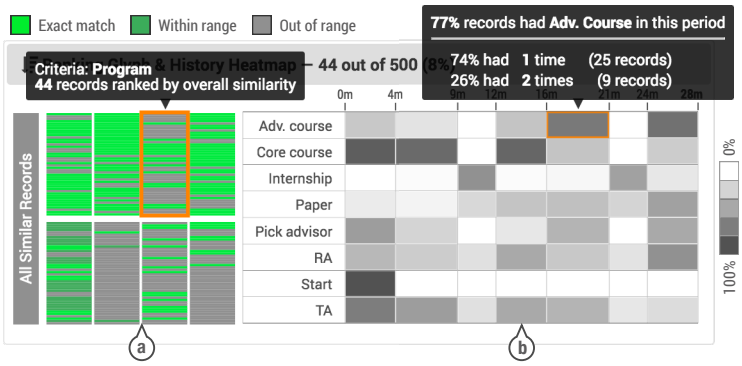

Fig. 5. (a) Ranking Glyph and (b) History Heatmap summarizing both criteria values and temporal activities of 44 most similar records. The figure includes two separate tooltips that would be shown when hovering on a glyph or a time period of the heatmap. In the Ranking Glyph, we see that the top portion of the highlighted "Program" glyph has few green bars. In comparison, for the "Paper & Advisor Pattern" glyph (second row, fourth column) most green matching records are at the top, indicating that the top records have the right pattern and that this criterion may have a strong influence on the overall similarity.

In summary, the LikeMeDonuts is a novel and highly customizable overview of a peer group that allows users to rapidly evaluate the similarities and differences of records in the group compared to the seed record. Interaction allows users to remove subsets directly in the LikeMeDonuts, spot matching controls or records in other coordinated views, and reorganize the rings.

3.3.4 Alternative Designs. Before settling on a sunburst-like circular layout, we explored alternative designs for presenting the similar records and the similarity criteria. We tested parallel coordinates [50] and radar plots [10], two common designs for visualizing multi-dimensional data. They were effective at revealing patterns between adjacent dimensions. However, since the dimensions are not hierarchically structured, it is difficult to track a group of records that share similar values across multiple criteria (e.g., male patients aged around 60 with Hyperglycemia) or to show the size of a group. Also, parallel coordinates have severe overlapping issues when displaying categorical values.

We also tested icicle plots and Treemaps [2], but as we compared all those designs our desire to center the design around the seed record (and a photo of the person) become stronger and we narrowed our design space to only circular designs. The classic sunburst design was enhanced and adapted to our application: (1) the hierarchy of similarity criteria can be reordered, (2) a set of operations allow users to modify the hierarchy and layout based on preferences, and (3) the photo at the center provides a visual reminder that all the information is relative to that person.

3.4 Ranking Glyph

The role of the Ranking Glyph is to help users understand how similarities and differences for each criterion evolve as they go down the ranked list of similar records. Each glyph represents a criterion and each horizontal bar within a glyph represents a record (Figure 5a). Records are ranked by their similarity to the seed record in all glyphs, with the most similar ones at the top and least similar ones at the bottom. The same consistent color scheme is applied. Bright green bars indicate that the criteria value of those records are identical to the value of the seed record while dark green bars represent records with criteria values within user-specified tolerance ranges. Records with criteria values outside tolerance ranges are shown as gray bars. The glyphs are arranged in the same layout

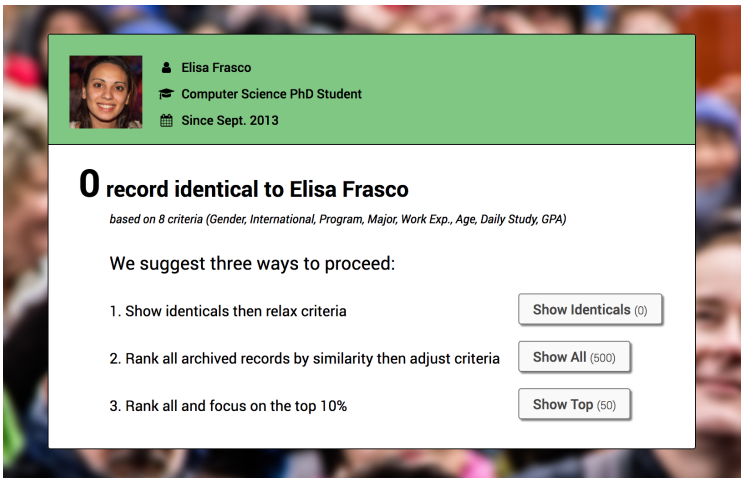

Fig. 6. The startup screen that shows basic information of the seed record and suggests three workflows for users to start the analysis: (1) show identical records, (2) show all archived records, and (3) show top 10% most similar records.

as the criteria controls (Figure 1b) and the record ranked list (Figure 1f). Hovering on a glyph highlights the focused criterion in other visualizations. Records selected in other visualizations will be highlighted in orange in the Ranking Glyph, revealing their positions in the ranked list.

3.5 History Heatmap

The History Heatmap summarizes the temporal events of the entire peer group or any selected subset of records. Each row of the timetable represents an event category and each column represents a time period (Figure 5b). In the example of students' academic records, each time period is a semester (e.g., Spring, Summer, and Fall). The darker the color of a cell the more events occurred in the time period, revealing hot spots in black (such as unsurprisingly"Start" in the first semester) and unpopular event in white (e.g., "Advanced Course" in Summer). When users select a subset of the similar records in other visualizations (e.g., by clicking on a ring sector in LikeMeDonuts), their activities will be shown in the history Heatmap, using an orange color gradient.

3.6 Support for Analytic Workflows

The first thing users typically do is to select the seed record. This would be done in the larger application in which PeerFinder may be embedded (e.g., EventAction [12]). Then the startup screen displays basic information about the seed record and provides a choice of three possible ways to proceed, i.e., three workflows[1], which come with explanations (Figure 6). These three workflows were born from the observations and interviews of users of the initial version of PeerFinder [13], and from our own discussion of possible ways to get started with the process.

The "Show Identicals" workflow helps users start with a small set of records and then relax constraints. It presents users with all similarity criteria set to "Exact Match" at the start and finds only identical records. Users can then adjust tolerance ranges or relax some of the criteria to "Close Match" to find a larger set of records with similar values. The second workflow is "Show All", which starts by selecting everybody and letting users review how the seed record differs from

[1]Video illustrating the interface and workflows is available in the supplemental material.

the entire population. All criteria are set by default to "Close Match" and all records are selected in the similarity distribution panel (Figure 2c). Users can then narrow the results by switching the criteria to "Exact Match", narrowing tolerance ranges and thus reducing the total number of records in the results. The "Show Top" workflow also uses "Close Match" for all criteria at the start but narrows the results to the top 10% most similar records. Users can further adjust the criteria and the similarity range to narrow or expand the results.

3.7 Interface Configuration Panel

On the top of the interface is a configuration panel that allows users to control the visibility of each interface component, so that different interface configurations can be used during different analysis stages. Pressing the "ESC" key will show and hide the interface configuration panel. Users can also rearrange the layout by drag-and-drop of interface components. In the usability study, we saw participants hide the criteria control panel and similarity distribution panel after they were happy with the criteria settings, and move the History Heatmap next to the seed record timeline to compare the activity patterns.

4 USER STUDY AND ITERATIVE DESIGN PROCESS

The new design evolved over three rounds of a formative usability study to evaluate the comprehensibility and learnability of the interface components and gain an understanding of users' analytic workflows in forming groups of similar people. We summarize the study procedure and report on users' feedback, describing how our prototype evolved.

4.1 Participants and Apparatus

We recruited a total of 12 university students by email (5 males and 7 females, aged 22–31, $M = 26.33$, $SD = 3.08$). All participants used computers in their study. The entire study was spread over three rounds during a month. In each round, we conducted study sessions with four participants and iteratively improved the prototype based on their feedback. A desktop computer was used, with a 24-inch display of resolution 1920×1200 pixels, a mouse, and a keyboard. Each participant received 10 dollars.

4.2 Dataset

We constructed a synthetic dataset of 500 archived records of university students with realistic but simplified features. The records had four categorical attributes: gender (male or female), major (Computer Science, HCI, Math, or Art), program (B.S., M.S., or Ph.D.), and international student (yes or no); four numerical attributes: age (when they started school), GPA, previous work experience (year), and average study time per day (hour). Eight categories of temporal events were included: "start school", "core course", "advanced course", "paper", "TA (teaching assistant)", "RA (research assistant)", "pick advisor", and "internship". On average, each archived record contained 35 events over 5 years. We generated record attributes with normal and binomial distributions. For temporal events, we reviewed real data and included similar patterns with random variations. The names of events and attributes are generic so that all students can conduct the tasks.

We handpicked one of the synthetic records to serve as the seed record. Her name is Elisa Frasco. The photo is authorized for using in mock-ups[2]. Elisa is an imaginary female international student, majoring in Computer Science and currently in the third year of her Ph.D. study. She is 24 years old and has one year of work experience before starting graduate school. On average, she spends 8 hours on study each day and maintains a GPA of 3.65. Her timeline shows no papers in the first two

[2] https://randomuser.me

years, internships in the last two summers, working as a TA all along except for an RA position in the last semester, after picking an advisor.

4.3 Procedure

Each session lasted about an hour. During the first five minutes, the experimenter made sure that participants were familiar with the task and the hypothetical friend. We told participants: *"You will be asked to (1) learn about a (hypothetical) close and important friend of yours who needs advice to improve her academic plan, such as when to take advanced classes, whether to intern during the summer, or when to try to publish papers, and (2) use a user interface to search for students similar to that friend. Data from those similar students will be used as evidence to provide guidance for your friend. You will not be asked to provide or review the guidance itself, only to select a set of similar students."* The record attributes and temporal events of the hypothetical friend were provided in a table and participants were encouraged to get familiar with it. Questions were answered.

Next, the startup screen of the interface was shown (Figure 6), and participants were encouraged to think aloud, explain the decisions they made, and comment on the interface. Participants decided what workflow option they wanted to use on their own and entered the main interface. No training was provided prior to the start. Participants explored the interface on their own and used the similarity criteria controls and visualizations to complete the task. If a participant was stuck for three minutes not being able to do what they wanted (e.g., did not understand an element of the interface) the experimenter provided hints and answered questions. The participants were reminded to care about their friend and there was no time limit for the task. The study session ended when the participant was satisfied with the peer group. If they had not used a component of the interface, the experimenter asked them to try it. At the end of the session, we asked participants to go back to the startup screen and try the other workflows. We collected learnability problems, comments, and suggestions for improvements.

4.4 Results and Evolution of the Design

We report on the participants' preferences toward the three workflows, and then focus on the three new visualizations. We report on users' feedback and describe how our prototype evolved.

4.4.1 Analytic Workflows. All participants seemed able to understand the workflow options provided on the startup screen on their own. The "Show Identicals" workflow was the most popular and was selected by seven out of 12 participants. "Show Top" and "Show All" workflows were used by three and two participants, respectively. Two participants in the first round complained that it was hard to anticipate the amount of data they would have to look at using the three options. We addressed this issue by adding the number of records next to the workflow options. After completing the task, we asked participants to try the other workflow options for 5 minutes each and rank the three options by preference. Three participants who had initially selected "Show Identicals" during the task changed their mind. Eventually, "Show Top" was the favorite of 5 participants, followed by "Show Identicals" (4) and "Show All" (3). Since there was no clear winner, we decided to keep all three workflows. In the future, usage logs from a larger number of users might help identify an adequate default workflow.

One common reason provided for favoring "Show All" and "Show Top" was wanting a larger number of similar records to get started. Comments included: *"it shows me a big picture"* and *"the overview helps me understand what I am dealing with."* In particular, one who preferred "Show Top" explained *"it starts with a good set of similar records and saves my time,"* and another said *"it guarantees some good results."* The participants who did not like "Show All" complained that showing all the data was overwhelming and one said *"I am lost. Seeing everything equals to seeing*

nothing." Another participant pointed out that *"the show all (workflow) is not scalable. It will destroy the visualizations and slow down the system."* Two participants were concerned about the biases in "Show Top" and explained: *"I want to see all the data instead of a small sample picked by the system."*

From the five participants who liked "Show Identicals" we heard comments such as: *"I preferred to start simple"* or *"the (ranking) algorithm was not involved and I had a better feeling of control."* However, others complained that *"it takes a longer time to get enough results"* and that *"start from blank was frustrating. I thought the system was broken."* Two participants pointed out that they would choose "Show Identicals" or "Show All," depending on the analysis, as one said: *"If I have a strong purpose such as predicting my job after graduation, I will start with only identicals and prepare queries based on my questions. Otherwise, I will start with all the data and try different (criteria) settings in a data-driven way."*

4.4.2 LikeMeDonuts. All participants were able to understand the meaning of the donut rings on their own. The color scheme was also understandable. One participant applauded that *"the color scheme is the same everywhere in the system. I learned it from the criteria controls."* Another said *"the color legend and the text labels made it clear to me."* Participants heavily used LikeMeDonuts just after they finished selecting the initial criteria settings. They mainly focused on reviewing the gray sectors and often went back to adjust the criteria controls to *"exclude unexpected records."* All participants left some gray sectors in the final results. One explained that *"I am aware about the gray areas but those criteria are less important. I will filter them out if I want fewer records in the results."* Another participant who deliberately balanced the gender of the peer group said: *"The gray records are not errors but expected. I kept the male students in gray to show the diversity."*

All participants commented positively about the four-stage animated transitions of LikeMeDonuts and everyone mentioned that the animations helped them keep track of the changes. We asked the participants to turn off the animations and explore for a few minutes. One participant was immediately confused and said aloud: *"Already? It updates too fast and I did not even notice."* Another pointed out that *"this is a complex interface. The animation helps me manage it."* However, later in the analysis, five participants changed their mind. One explained *"the animations take time to play and slow down my operations."* Another added: *"As I become familiar with and trust the system, I may want to turn it off."* So providing the option to turn off the animation is required. At the end, all participants strongly agreed that animations are important for new users to learn the system, confirming previous findings (e.g., [39]). Seven participants stated that they will keep the animation active all the time. One said *"it does not take that much time"* and another emphasized *"I make mistakes sometimes. It helps me verify my operations."*

As for sorting the donut rings, nine out of 12 participants moved important criteria to the inner rings and kept less important ones in the outer rings. One explained that *"I read the donuts from inside to outside"* and another said *"I prefer to keep important things around my friend."* Two participants used the opposite order because *"the outer rings have more space for important criteria."* The last participant used a mixed strategy. He first sorted the criteria by the number of unique values and then by their importance. During the first round of the usability study, no participant had discovered that they could reorder the donut rings by dragging the criteria icons. We improved this by adding a dragging handler to the icons and changed the mouse cursor to a "Move" style when hovering on the icon. This helped all remaining participants discover the feature.

In the first two rounds of the study, three out of 8 participants had not been able to understand the ring sectors on their own. Two blamed the grouping of sectors in the inner rings: *"I see only a few divisions in the inner rings but many in the outer rings. I did not realize that the rings are aligned to show individual records."* They suggested two ideas to improve learnability: (1) removing the grouping and drawing borders to separate individual records, and (2) highlighting individual records

with borders when hovering on a sector. We implemented the second solution and kept the grouping, which helps LikeMeDonuts scale to larger numbers of records. The four remaining participants discovered the meaning of the donuts on their own and commented that the highlighting was helpful for reviewing individual records.

The thin partial ring sectors outside the donuts were not available during the first two rounds of the study. We observed four out of 8 participants pointing fingers at the screen trying to identify individual records with all criteria values in green. One of them explained that *"I wanted to see if there were any identical records after I changed the criteria."* We then designed the thin ring to highlight identical records and all four participants in the third round were able to understand it on their own. *"I found the green arc when I was focusing on a very similar record,"* one commented, *"I immediately understood."*

4.4.3 Ranking Glyph. The initial design of the Ranking Glyph came from brainstorming ideas to represent how the top records differed from the bottom records. In the initial prototype, we had placed the Ranking Glyph on the left side, below the criteria control panel. We hoped that users would understand the Ranking Glyph layout from the layout of the criteria. However, none of the participants of the first round of testing guessed the meaning of the glyph. After being explained how the Ranking Glyph worked one participant said that *"it (the glyph) looks like a compressed version of the records"* and that *"they are both sorted by similarity,"* suggesting that the Ranking Glyph should be moved next to the similar record ranked list (Figure 1f). We moved the Ranking Glyph to the top of the record list and we also moved the History Heatmap next to it (previously displayed as the background colors of the seed record timeline). These two visualizations combined provide a true overview of the results.

In the subsequent two rounds of usability testing, all participants were able to understand the Ranking Glyph on their own. The most common learning strategy was to look at individual records first in the ranked list. *"It looks like a barcode of the record list"* one participant commented. Further testing should verify that the glyph is still learnable when the individual records are hidden for privacy reasons.

The participants typically used Ranking Glyph to determine a similarity threshold, as one said: *"I used the barcharts to filter by value and used the glyphs to filter by similarity."* Five participants particularly liked the way the glyphs are sorted. One said *"it is useful for checking how the top 5% and bottom 5% records look like."* Another (who had some data mining background) commented: *"The glyphs can tell me how each criterion influences the overall similarity. I can easily see the trivial (less influential) ones and put less weight on them."*

The main complaint about the Ranking Glyph is its small size. One participant complained that *"it is too small and hard to track individuals. I can see the top 5% records but cannot see the fifth record."* Another said *"I am more willing to interact with the donuts than the ranking glyph. The pixels (bars) are too small."* To mitigate this issue, we connected Ranking Glyph with LikeMeDonuts so that users benefit from both the interactivity of the donuts and the sorted overview of the Ranking Glyph. Specifically, when a subset of records are selected in the donuts, the horizontal bars representing those records are highlighted, showing their rankings in the entire peer group (Figure 4c).

4.4.4 History Heatmap. In the first round of the study, the history heatmap was on the left side, combined with the seed record timeline (displayed as background color of the seed record event squares in each cell). The first-round participants were not able to tell if it was showing the activities of all archived records or similar records. Therefore, we moved the History Heatmap to the top of the similar record ranked list, right next to the Ranking Glyph. All remaining participants were able to guess the meaning of the color darkness in the History Heatmap on their own. One

participant stated: *"The heatmap is intuitive, just like you add up the gray squares in the timelines below."*

Six participants reported findings from the History Heatmap. For example, *"I was able to see the transition from core course to advanced course and hotspots of internships in the summer,"* one said, *"some interesting patterns just jump into my eyes."* We also observed two participants using the History Heatmap to help understand the activity of the seed record, as one explained *"I wonder if my friend has done any abnormal thing."* To better support this task, we now allow users to change the layout of the interface components so the History Heatmap can be moved closer to the seed record timeline to compare the activities side-by-side. In the future, adding a way to compute the difference between two records or between the seed record and the peer group average may be useful for the timeline views.

4.4.5 Similar Record Barcharts. All participants were able to understand the similar record distribution barcharts (Figure 4b) immediately and could correctly tell the meaning of colors, horizontal axis, and heights. During the analysis, the participants typically used the barcharts to briefly review the criteria distributions when adjusting the criteria controls. One participant explained that *"it helps me verify my settings"* and another added that *"it just looks simpler than other visualizations."*

When asked to compare barcharts to LikeMeDonuts, all participants preferred LikeMeDonuts. Three reasons were commonly mentioned. First, LikeMeDonuts provides the capability to track individual records (by following a radius of the circle), which is not possible in barcharts. *"The donuts show an extra level of information"* one participant explained. Second, LikeMeDonuts shows an overview of the entire peer group while barcharts only show overviews of individual criteria. One participant stated that *"when I turn off the labels and step back, I can estimate the overall similarity of the group from the colors"* and another commented that *"using barcharts, I need to read eight separate charts. I only need to focus on one chart using the donuts."* In contrast, they thought barcharts were only useful for reviewing a single criterion at a time, as one participant said: *"It makes no sense to compare the bars between two criteria, like the number of female students to the number of computer science students."* Finally, five participants mentioned that LikeMeDonuts is more aesthetic and one added: *"It looks cool. I feel more motivated to show this to my friend."*

The participants also pointed out that a unique advantage of barcharts is that they make visible trends in the criteria values, e.g., *"it shows me the overall shape and I can clearly see records with extreme criteria values,"* or *"barcharts can guide me to filter out outliers."* In comparison, they found it difficult to review criteria distributions in LikeMeDonuts, where criteria values are repeatedly split within each branch. *"Values in the outer rings are not aggregated and I need to review the sectors one by one,"* commented by a participant during the second round of study. To address this weakness, we coordinated LikeMeDonuts with barcharts: when users click on a subgroup in the donuts, the distributions of the selected records will be highlighted in the barcharts (Figure 4b). This enables users to review the criteria distributions of subsets of records.

5 CASE STUDY

To evaluate the usefulness of the system with the new designs, we conducted a case study with a student review manager who has access to all student records. This person was a professor with 12 years of experience in advising graduate students in computer science. The case study took place over two weeks. During the first week, we demonstrated our prototype using a synthetic dataset. The review manager prepared a dataset of real students' data. During the second week, we deployed our prototype to the review manager's workstation and he used our prototype to perform the analysis. We recorded the review manager's analysis process, findings, and feedback.

5.1 Data Preparation

The review manager prepared a dataset of 641 archived records of graduate students in the computer science department. The dataset consists of students' temporal activities including courses (core or advanced), assistantships (teaching or research), publications, and milestones (start school, done classes, and advance to candidacy). Students' record attributes include numbers of grades (As, Bs, and Cs), numbers of assistantships (teaching and research), number of publications, class status (coursework completed or not), and candidacy status (advanced or not).

During the analysis the data from one of the authors of the paper (a fourth-year Ph.D. student) was used as the seed record. All other records were de-identified. The analysis goal was to find a group of students similar to him, so that follow-up analyses may be conducted based on the similar records, such as predicting the first placement of the seed record after graduation and generating recommendations to help the seed record make academic plans for the next year.

5.2 Reviewing All Data

The review manager started with the "Show All" workflow to obtain a complete overview of the entire data. After about 5 seconds, the data were loaded and visualizations rendered. The review manager first explored the barcharts to inspect the criteria distributions of all archived students. He verified that the criteria values matched his expectations, for example, the percentages of students who had done classes and who had advanced to candidacy, and the distribution of the course grades.

Then, the review manager explored the record timelines and History Heatmap to review temporal information. He first looked at the seed record's history activities during the last four years and found 8 consecutive research assistantships since year one. *"Your research assistantship started early,"* he commented, *"this could be a useful pattern."* The review manager also noticed two B grades during the seed record's second year of study. He specified these two temporal patterns as similarity criteria using the seed record timeline panel.

He then reviewed the temporal activities of all archived students. The History Heatmap showed an activity summary and confirmed several of his expectations, e.g., that there is a transition from teaching assistantship to research assistantship starting in the third semester, and that most students achieved the "done with classes" milestone between the third and the sixth semester as required by the department. However, two findings were unexpected. First, the students started to receive fewer As in the third semester. The review manager thought this could be caused by the increase in the difficulty of the advanced courses, or due to the fact that a number of students had finished taking classes and thus no grades were recorded. Second, nearly twice as many publication events occurred in the Spring semester than in the Fall semester. The review manager was unsure about this phenomenon. One hypothesis may be that many conferences in computer science announce paper acceptances in the Spring and hold the conference later in the year.

The review manager then explored the similarity score distribution. He found that the shape of the distribution had two peaks, where the first peak contained the top 37% most similar records, and the second peak was taller and contained the remaining records. *"This looks strange,"* he said, *"I was expecting a normal distribution with one peak."* To understand how the peaks were formed, he selected records in the second peak. By looking at the barcharts, he realized that those are all new graduate students in their first or second year: they all had less than four assistantships, had not yet finished classes or advanced to candidacy. The review managed said: *"Now it makes sense. The first peak are senior students like you and the second peak are junior students unlike you. We only need those senior ones."* He then selected the top 10% most similar students and started using other visualizations to review in detail.

5.3 Reviewing Similar Records

The review manager started reviewing similar records using LikeMeDonuts. Immediately, he found that only a few exact matches were bright green while most of the sectors in the donuts were gray. *"The colors help me estimate the overall quality of the peer group,"* he commented, *"I will add some tolerance and try to make it about 50% green before reviewing in detail."* As he was adjusting the tolerance ranges, he noticed a unique branch of records in LikeMeDonuts: these records were included as the top 10% most similar but they had not done classes yet. The review manager followed the branch to inspect other criteria values of them. *"This is weird,"* he said after exploring for a while, *"they all have advanced to candidacy but not done classes. We may have errors in the data."* He clicked on the branch of LikeMeDonuts and records were highlighted in the barcharts, Ranking Glyph, and record ranked list. He reviewed the record ranked list to check the temporal activities. The review manager found that those students did not even have start school milestone events. He suddenly realized that they were probably transfer students brought in by professors who moved to the university. *"Their candidacy status was transferred to our department but some of their courses were unqualified to transfer,"* he explained, *"we may leave them in the results but keep the gray color to be noticeable."*

The review manager then explored the Ranking Glyph. He read the glyphs one by one and found three types of patterns: (1) green on the top and gray on the bottom (e.g., research assistantship and publications), (2) dominated by green or gray (e.g., done classes, advanced to candidacy), and (3) alternating between green and gray (e.g., course grades). *"Some criteria seem more correlated to the overall similarity and have a larger impact on the ranking,"* he commented and adjusted the criteria controls to increase the weights of research assistantship and publications, and reduced the weights of course grades. *"The alternating pattern indicates that individual course grades are not good features to characterize graduate students,"* he added.

5.4 Summary and Feedback

Overall, the review manager found the prototype very effective for finding similar students and enable a data-driven way for student advising. When asked about his preferences for the visualizations and analytic workflows, he stated that *"the three visualizations all have their own uses that cannot be easily replaced by each other."* He expressed some enthusiasm for the Ranking Glyph because *"it provides an effective overview to understand the effect of each criterion on the overall ranking."* He also liked the use of color in LikeMeDonuts because it *"provides a good overview of the quality of the similar records."*

The review manager stated that he preferred to use the "Show All" workflow, especially when working with a new dataset: *"Starting with all the available data helps obtain an unbiased overview and provides means to check the data quality and discover initial findings to guide the analysis."* He also emphasized that *"understanding why those least similar students are different from you can also provide insights."* However, he expressed concerns about the visual clutters and interaction latency when showing all the data as the number of records becomes extremely large. In the end, the review manager applauded that *"seeing both attributes and temporal activities is important for reviewing student records. I appreciate that your system provides visualizations for this purpose."*

6 DISCUSSION

We discuss the limitations and new opportunities discovered in our study.

6.1 Limitations

All the study participants were university students, so a more diverse population should be tested to further improve the interface. Larger numbers of participants and longer periods of use may alter usage patterns and lead to new strategies and other analysis workflows. The interface can be further improved. For example, the green similarity color encoding could be applied to the timelines as well and missing data may have to be represented with a separate encoding.

Scalability becomes an issue for most interactive visualizations as the size of the data grows. Our system prototype runs smoothly with a testing dataset of 10,000 records, each with an average of 40 events. A larger number of archived records can slow down the computation of similarity and the rendering of the visualizations. Better techniques to cluster and compare records in groups would enhance the performance for applications requiring extremely large datasets, such as millions of online customer records. When the number of criteria grows larger, showing all criteria at once is likely to overwhelm most users (as illustrated in Figure 7). Automatically selecting two or three criteria to start may be useful [33, 53]. Splitting the criteria into multiple LikeMeDonuts may also be useful (e.g., one for demographics, another for academic experience, and a third for work experience), but evaluation is needed to identify and quantify benefits, and other solutions may emerge.

Applying the interface to other application domains is likely to reveal further issues. For example, we know that more advanced temporal query methods [27, 34] will need to be integrated to tackle most medical applications. Other data types need to be supported, e.g., network connections between individuals [3, 57, 60]. Our study mainly focused on the scenario of making important life decisions when users demand more controls and context even at the cost of added complexity [13, 23]. Our designs and findings may not be applicable to recommender systems for making less important decisions in entertainment and shopping applications.

Finally, while most students, patients, and others who must make life choices are eager to follow the paths of predecessors, there are dangers to such an approach. Biases may be introduced when the data available do not represent people adequately or when there are errors or missing attributes in the data [36]. Decision-makers who consult databases of predecessors risk repeating old paths which are no longer relevant because past histories of bias have been rectified or because circumstances have changed. While there may still be lessons from the past, users need to be reminded that their history is unique and that breaking from past paths may be a powerful way to distinguish themselves. Visual analytics solutions may already be a big improvement compared to black box solutions, but how do we provide guard rails to limit the effect of possible biases?

6.2 New Opportunities

While automated black-box recommendation techniques are effective and used widely in entertainment and shopping applications, transparency is critical when users review recommendations for important life decisions [23, 44]. Our early investigation suggests that visual representations such as the LikeMeDonuts can help users review similarities and differences in the peer student group. Another example with a real dataset of professors is shown in Figure 8.

Beyond similarities and differences, "DiverseDonuts" can also be designed to guide the creation of diverse teams. Diversity can drive innovation in teams [24]. An organization may need to assemble a panel of peers to review the grievance brought up by an employee. In this case, the group of peers needs to be close to the employee but diverse enough to include members from diverse divisions of the company, genders, backgrounds, and with some age and background variations. Detecting clusters and selecting representative records from each cluster is a potential approach to pursue.

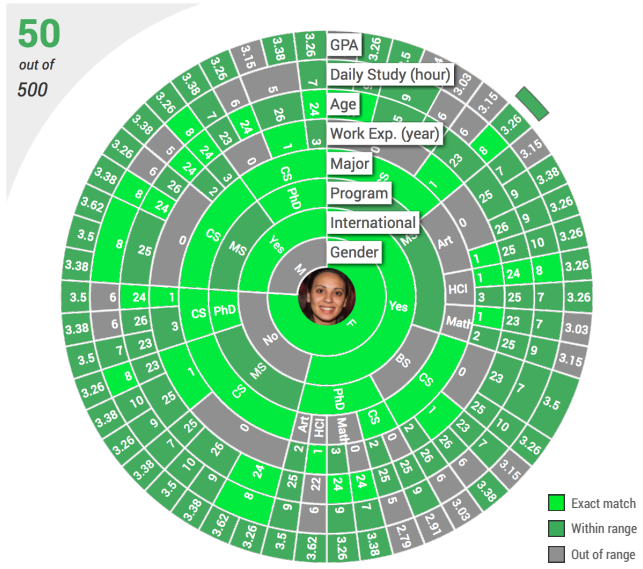

Fig. 7. The LikeMeDonuts showing all the 8 criteria of the student dataset used in the usability study.

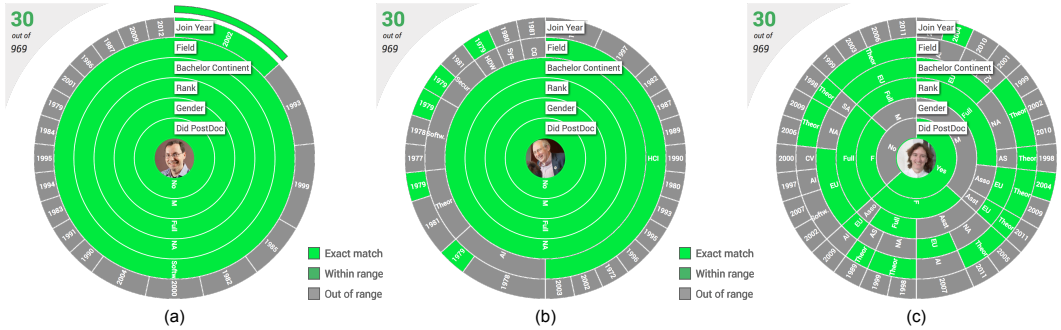

Fig. 8. This example uses a real dataset of 969 professors from top Computer Science Graduate Programs [37]. The three LikeMeDonuts visualizations show the top 30 most similar records of (a) Dr. David M. Brooks, (b) Dr. Ben Shneiderman, and (c) Dr. Claire Mathieu. The most similar records of Dr. Brooks are all identical to each other except for joining year. Dr. Shneiderman is unique in research field and joining year compared to his peer group but normal in other criteria. The peer group of Dr. Mathieu is very diverse for all criteria.

Finally, we believe that tools such as the one described in this paper can help data scientists define better distance metrics that can then be used automatically in some situations after proper evaluations are conducted.

7 CONCLUSION

Recommendation applications can guide users in making important life choices by referring to the activities of similar peers. In this paper, we focus on how to improve the selection of peer groups. We have described a novel set of visual techniques (LikeMeDonuts, Ranking Glyph, and History Heatmap) and a visual encoding of similarity, which can be combined with basic methods for criteria selection and timeline views. The resulting combination and user-controlled selection

of workflows enable users to rapidly evaluate the similarities and differences in a peer group compared to a seed record. Interaction facilitates the review of aggregated summaries as well as individual record views and their ranking. A formative lab evaluation and a case study with real data strengthen our belief that finding "people like me" is a challenging problem that will greatly benefit from visual analytics approaches. While similarity between people will remain a subjective measure and vary based on the context of use, the creation of ground truth datasets for specific situations will pave the way to more formal evaluation.

ACKNOWLEDGMENTS

The authors would like to thank all the participants involved in the studies and the reviewers for their valuable feedback. This work was supported in part by Adobe Research.

REFERENCES

[1] F Gregory Ashby and Daniel M Ennis. 2007. Similarity measures. *Scholarpedia* 2, 12 (2007), 4116.
[2] Benjamin B Bederson, Ben Shneiderman, and Martin Wattenberg. 2002. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Transactions on Graphics* 21, 4 (2002), 833–854.
[3] Nan Cao, Yu-Ru Lin, Fan Du, and Dashun Wang. 2016. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications* 36, 5 (2016), 72–81.
[4] William B. Cavnar and John M. Trenkle. 1994. N-gram-based text categorization. In *Proceedings of the Annual Symposium on Document Analysis and Information Retrieval*. 161–175.
[5] Florin Chelaru, Llewellyn Smith, Naomi Goldstein, and Héctor Corrada Bravo. 2014. Epiviz: Interactive visual analytics for functional genomics data. *Nature Methods* 11, 9 (2014), 938–940.
[6] Fanny Chevalier, Nathalie Henry Riche, Catherine Plaisant, Amira Chalbi, and Christophe Hurter. 2016. Animations 25 years later: New roles and opportunities. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 280–287.
[7] Richard Chimera. 1992. Value bars: An information visualization and navigation tool for multi-attribute listings. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 293–294.
[8] Gouthami Chintalapani, Catherine Plaisant, and Ben Shneiderman. 2004. Extending the utility of treemaps with flexible hierarchy. In *Information Visualisation*. 335–344.
[9] Lieven Decock and Igor Douven. 2011. Similarity after goodman. *Review of Philosophy and Psychology* 2, 1 (2011), 61–75.
[10] Geoffrey M Draper, Yarden Livnat, and Richard F Riesenfeld. 2009. A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 15, 5 (2009), 759–776.
[11] Steven Drucker and Roland Fernandez. 2015. *A Unifying Framework for Animated and Interactive Unit Visualizations*. Technical Report. Microsoft Research.
[12] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2016. EventAction: Visual analytics for temporal event sequence recommendation. In *Proceedings of the IEEE Visual Analytics Science and Technology*. 61–70.
[13] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2017. Finding Similar People to Guide Life Choices: Challenge, Design, and Evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 5498–5509.
[14] Niklas Elmqvist and Jean-Daniel Fekete. 2010. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 439–454.
[15] Lyndsey Franklin, Catherine Plaisant, Kazi Minhazur Rahman, and Ben Shneiderman. 2016. TreatmentExplorer: An interactive decision aid for medical risk communication and treatment exploration. *Interacting with Computers* 28, 3 (2016), 238–252.
[16] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309.
[17] Michael Glueck, Alina Gvozdik, Fanny Chevalier, Azam Khan, Michael Brudno, and Daniel Wigdor. 2017. PhenoStacks: Cross-sectional cohort phenotype comparison visualizations. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 191–200.
[18] Michael Glueck, Peter Hamilton, Fanny Chevalier, Simon Breslav, Azam Khan, Daniel Wigdor, and Michael Brudno. 2016. PhenoBlocks: Phenotype comparison visualizations. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 101–110.
[19] Carlos A Gomez-Uribe and Neil Hunt. 2016. The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems* 6, 4 (2016), 13:1–13:19.

[20] Fabio Grandi. 2010. T-SPARQL: A TSQL2-like temporal query language for RDF. In *In International Workshop on Querying Graph Structured Data*. 21–30.
[21] Marti A Hearst. 1995. TileBars: Visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 59–66.
[22] Jeffrey Heer and George Robertson. 2007. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1240–1247.
[23] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 241–250.
[24] Sylvia Ann Hewlett, Melinda Marshall, and Laura Sherbin. 2013. How diversity can drive innovation. *Harvard Business Review* 91, 12 (2013), 30–30.
[25] Harry Hochheiser and Ben Shneiderman. 2004. Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization* 3, 1 (2004), 1–18.
[26] Daniel A Keim. 2000. Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics* 6, 1 (2000), 59–78.
[27] Josua Krause, Adam Perer, and Harry Stavropoulos. 2016. Supporting iterative cohort construction with visual temporal queries. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 91–100.
[28] Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 6822 (2001), 860–921.
[29] Weizhong Li and Adam Godzik. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 13 (2006), 1658–1659.
[30] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing* 7, 1 (2003), 76–80.
[31] Ran Liu, Tiffany Chao, Catherine Plaisant, and Ben Shneiderman. 2012. ManyLists: Product comparison tool using spatial layouts with animated transitions. *University of Maryland Technical Report* (2012).
[32] Sana Malik, Ben Shneiderman, Fan Du, Catherine Plaisant, and Margret Bjarnadottir. 2016. High-volume hypothesis testing: Systematic exploration of event sequence comparisons. *ACM Transactions on Interactive Intelligent Systems* 6, 1 (2016), 9:1–9:23.
[33] Matthew Louis Mauriello, Ben Shneiderman, Fan Du, Sana Malik, and Catherine Plaisant. 2016. Simplifying overviews of temporal event sequences. In *CHI Extended Abstracts on Human Factors in Computing Systems*. 2217–2224.
[34] Megan Monroe, Rongjian Lan, Juan Morales del Olmo, Ben Shneiderman, Catherine Plaisant, and Jeff Millstein. 2013. The challenges of specifying intervals and absences in temporal queries: A graphical language approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2349–2358.
[35] Gonzalo Navarro. 2001. A guided tour to approximate string matching. *Comput. Surveys* 33, 1 (2001), 31–88.
[36] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.
[37] Alexandra Papoutsaki, Hua Guo, Danae Metaxa-Kakavouli, Connor Gramazio, Jeff Rasley, Wenting Xie, Guan Wang, and Jeff Huang. 2015. Crowdsourcing from scratch: A pragmatic experiment in data collection by novice requesters. In *AAAI Conference on Human Computation and Crowdsourcing*.
[38] Catherine Plaisant, Ben Shneiderman, and Rich Mushlin. 1998. An information architecture to support the visualization of personal histories. *Information Processing & Management* 34, 5 (1998), 581–597.
[39] Catherine Plaisant, Johnny Wu, A Zach Hettinger, Seth Powsner, and Ben Shneiderman. 2015. Novel user interface design for medication reconciliation: An evaluation of Twinlist. *Journal of the American Medical Informatics Association* 22, 2 (2015), 340–349.
[40] Katie Powell, John Wilcox, Angie Clonan, Paul Bissell, Louise Preston, Marian Peacock, and Michelle Holdsworth. 2015. The role of social networks in the development of overweight and obesity among adults: A scoping review. *BMC public health* 15, 1 (2015), 996.
[41] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 175–186.
[42] Simone Santini and Ramesh Jain. 1999. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 9 (1999), 871–883.
[43] Ben Shneiderman. 1994. Dynamic queries for visual information seeking. *IEEE Software* 11, 6 (1994), 70–77.
[44] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI Extended Abstracts on Human Factors in Computing Systems*. 830–831.
[45] Mike Sips, Jörn Schneidewind, Daniel A Keim, and Heidrun Schumann. 2006. Scalable pixel-based visual interfaces: Challenges and solutions. In *Information Visualization*. 32–38.

[46] Richard Snodgrass. 1987. The temporal query language TQuel. *ACM Transactions on Database Systems* 12, 2 (1987), 247–298.
[47] Michael Spenke. 2001. Visualization and interactive analysis of blood parameters with InfoZoom. *Artificial Intelligence in Medicine* 22, 2 (2001), 159–172.
[48] Melanie Swan. 2012. Crowdsourced health research studies: An important emerging complement to clinical trials in the public health research ecosystem. *Journal of Medical Internet Research* 14, 2 (2012), e46.
[49] Daniel Tunkelang. 2009. Faceted search. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, 1 (2009), 1–80.
[50] Junpeng Wang, Xiaotong Liu, Han-Wei Shen, and Guang Lin. 2017. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 81–90.
[51] Paul Wicks, Michael Massagli, Jeana Frost, Catherine Brownstein, Sally Okun, Timothy Vaughan, Richard Bradley, and James Heywood. 2010. Sharing health data for better outcomes on PatientsLikeMe. *Journal of Medical Internet Research* 12, 2 (2010), e19.
[52] Max L Wilson, Paul André, et al. 2008. Backward highlighting: Enhancing faceted search. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 235–238.
[53] Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2017. Voyager 2: Augmenting visual analysis with partial view specifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2648–2659.
[54] Krist Wongsuphasawat and Ben Shneiderman. 2009. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *IEEE Symposium on Visual Analytics Science and Technology*. 27–34.
[55] World Health Organization. 1992. The tenth revision of the international classification of diseases and related health problems (ICD-10). (1992).
[56] Yingcai Wu, Nan Cao, David Gotz, Yap-Peng Tan, and Daniel A Keim. 2016. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia* 18, 11 (2016), 2135–2148.
[57] Yanhong Wu, Naveen Pitipornvivat, Jian Zhao, Sixiao Yang, Guowei Huang, and Huamin Qu. 2016. egoslider: Visual analysis of egocentric network evolution. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 260–269.
[58] Mehmet Adil Yalçın, Niklas Elmqvist, and Benjamin B Bederson. 2017. Keshif: Rapid and expressive tabular data exploration for novices. *IEEE Transactions on Visualization and Computer Graphics* PP, 99 (2017), 1–14.
[59] Emanuel Zgraggen, Steven M. Drucker, Danyel Fisher, and Robert DeLine. 2015. (s|qu)eries: Visual regular expressions for querying and exploring event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2683–2692.
[60] Jian Zhao, Michael Glueck, Fanny Chevalier, Yanhong Wu, and Azam Khan. 2016. Egocentric analysis of dynamic networks with EgoLines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 5003–5014.
[61] Jian Zhao, Zhicheng Liu, Mira Dontcheva, Aaron Hertzmann, and Alan Wilson. 2015. MatrixWave: Visual comparison of event sequence data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 259–268.

Received August 2017; revised November 2017; accepted January 2018