ScholarLens Introduction

ScholarLens is a visual analytic tool for understanding the academic performance of a scholar or comparing the performance of multiple scholars. At the core of our design is the concept of **paper sets** and the operations that combine or subdivide the paper sets. For example, two primary components of one's scholarly performance: **productivity** and **impact**, are represented, respectively, by the paper set of her **publications** and the paper set of the **citations** received. Other factors, including independence, collaboration, and performance over time, can be demonstrated by the operations on the paper sets as well. For example, a scholar's ability to conduct research independently without the help from her advisor may be revealed her paper set – her advisor's paper set. This introduction will cover the detailed usage of ScholarLens' visual interface and interaction.
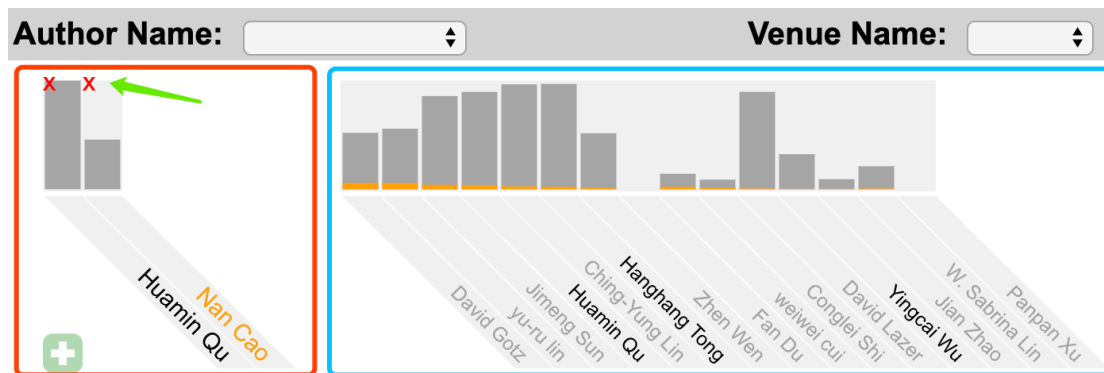
1. Adding scholars

Figure 1. The interface for adding scholars for investigation. The red and blue rectangles highlight the selected list and the co-author list, respectively.

A scholar can be added by *selecting her name from the dropdown box*. Once added, her name will be displayed in the **selected list** (as highlighted in the red rectangle), and her co-authors will be displayed in the **co-author list** (as highlighted in the blue rectangle). The height of a gray bar represents the number of publications corresponding to that scholar, and the height of the orange bar represents the number of co-authored papers. You may click on an author's name in the selected list to update her co-author list. Once switched, her name will be highlighted in orange. For example, Nan Cao is highlighted in Figure 1, indicating his co-author list is currently displayed. A scholar in the co-author list may be added to the selected list by *clicking her name in the co-author list*. You can remove the scholar from the selected list by clicking on the red "×" button (highlighted by the green arrow in Figure 1.)

2. Adding paper sets for selection

Figure 2. The interface for composing and adding a new paper set.

By clicking the "+" button in the bottom-left corner of the selected list, the interface for composing and adding a new paper set will be shown. In the interface, each column indicates the available operators for the corresponding scholar, and each row corresponds to the same operator for all selected scholars. The currently selected operator is highlighted in orange and displayed in the top row (above the horizontal line). For example, the selection in Figure 2 indicates that the "and", "or", and "or" operators are used for Huamin Qu, Nan Cao, and Yingcai Wu, respectively. The meaning of each operator is as follows:

- "**Not**" indicates *subtraction* operation, meaning that the papers authored by this scholar will be removed from the resulting paper set.
- "**Ignore**" indicates *no* operation will be applied, meaning that the scholar will not be considered when composing the paper set.
- "**And**" indicates *intersection* operation, meaning that any paper in the resulting set *must* be (co)authored by this scholar.
- "**Or**" indicates *union* operation, meaning that a paper in the resulting set *may* be (co)authored by any scholar with the "Or" operator.

For example, the selection in Figure 2 indicates the formula of "(Nan Cao | Yingcai Wu) + Huamin Qu". The meaning of the selection will be textually displayed on the right side of the selection as well. Once the selection is completed, you may click on the small green "+" button (highlighted by the red arrow in Figure 2) to add the paper set for further investigation.
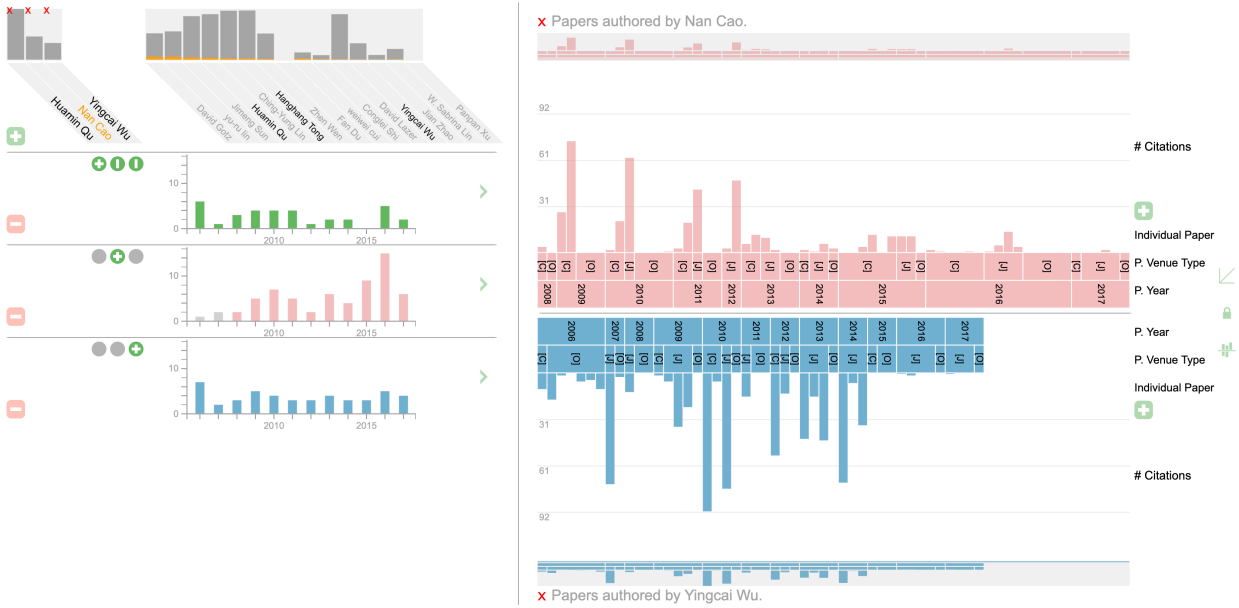
3. Selecting paper sets for investigation

Figure 3. The interface for selecting paper sets and adding them to the hierarchical histogram.

The previously composed paper sets are displayed under the selected and co-author lists. The selection of operators is displayed on the left side of each paper set to indicate how the paper set is composed. A bar in the histogram shows the number of publications of this paper set in a specific year. You may brush a time range for investigation. By default, the entire time range is considered. You may *click on the green arrow* on the right side of the histogram to add the paper set within the selected time range for detailed investigation with the hierarchical histogram, which is shown on the right side of Figure 3. The color of the selected bars will be updated using the color of the hierarchical histogram to build mental connections, and the meaning of the paper set will be displayed above/below the upper/lower histograms as well. Up to two paper sets can be added at the same time for comparison in the hierarchical histogram. One will be mapped to the *upper* histogram and the other to the *lower* histogram. The upper/lower histogram may be removed by clicking the red "×" on the left side of the title.

The hierarchical histogram features a **multilevel design**: *several* levels of *horizontal* bars at the *intermediate* levels of the hierarchy where the *width* of each bar indicates the corresponding number of leaf nodes (for Figure 3, the number of individual papers), followed by *one* level of *vertical* bars at the *finest* (i.e., leaf) level of the hierarchy where the *height* of each bar can indicate the number of papers, the total number of citations received by the papers, or the *h*-index of the papers.
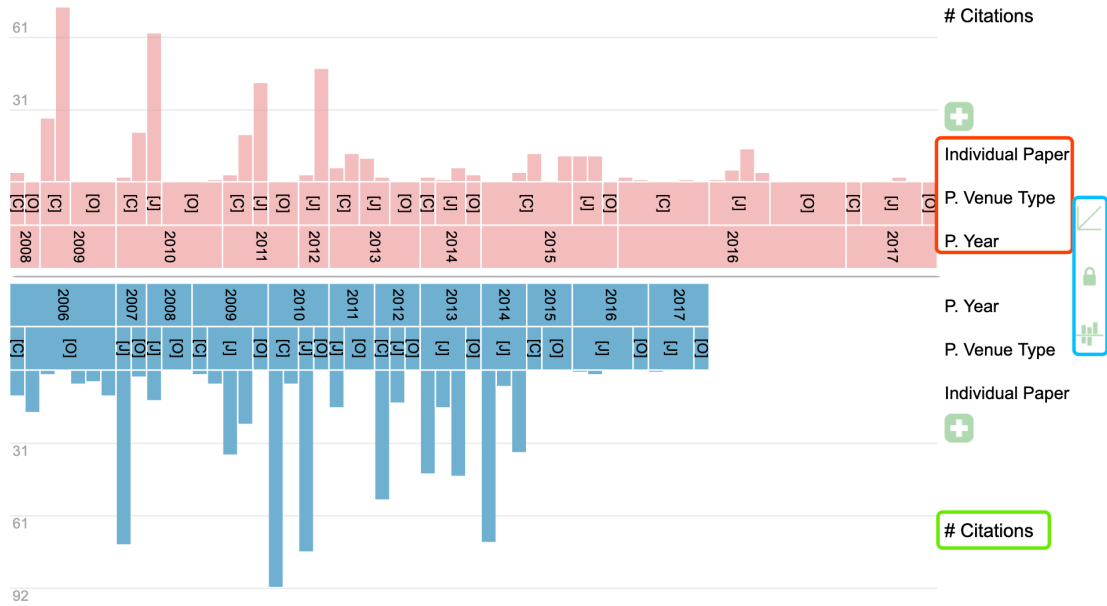
4. Operating the hierarchical histogram

Figure 4. The interface of hierarchical histogram.

Producing the hierarchy. The hierarchical histogram partitions the entire paper set into smaller groups in several levels. The grouping in each level is determined by an attribute, as highlighted in the red rectangle on the right of Figure 4. An attribute starting with "**P.**" indicates a *publication* attribute, and an attribute starting with "**C.**" indicates a *citation* attribute. For the example shown in Figure 4, both paper sets are partitioned by three attributes. First, the entire paper set is partitioned by the publication year (P. Year). Then, the paper subset in each year is further partitioned by the publication venue type (P. Venue Type, [C] Conference, [J] Journal, or [O] Others). Finally, each resulting subset is partitioned into individual papers (Individual Paper). You can **change** or **delete** an attribute by mouseover that attribute and selecting the corresponding item showing in the pop-up menu. More attributes may be added by clicking the green "+" button above (for the upper histogram) or below (for the lower histogram) the corresponding attribute list. You can swap the **partitioning order** of the paper set by *dragging an attribute up and down*.

Determining the height of the vertical bars. While each vertical bar has the same width, the height of the vertical bar can indicate one of the following attributes:
- "**# papers**" indicates the number of papers in the corresponding subset.
- "**# citations**" indicates the number of citations in the corresponding subset.
- "**h-index**" indicates the *h*-index of the corresponding subsets (i.e., the largest number of papers *h* that received at least *h* citations).

The default attribute is "# citations", and other attributes can be selected from the pop-up menu when mouseover the currently selected attribute (highlighted in the green rectangle).

Other functions. Three functions are provided to enhance usability, as highlighted in the blue rectangles. From top to bottom, these functions are:

- "**Bar Scale**": the height of the vertical bars at the ***finest*** (i.e., leaf) level of the hierarchy can be displayed in the *linear*, *square root*, or *logarithmic* scale. You may click on the scale button to switch among them.

- "**Attribute lock**": the attributes to partition the paper sets can be synchronized for the upper and lower histograms, once they are "locked"; otherwise, the attributes can be selected independently.
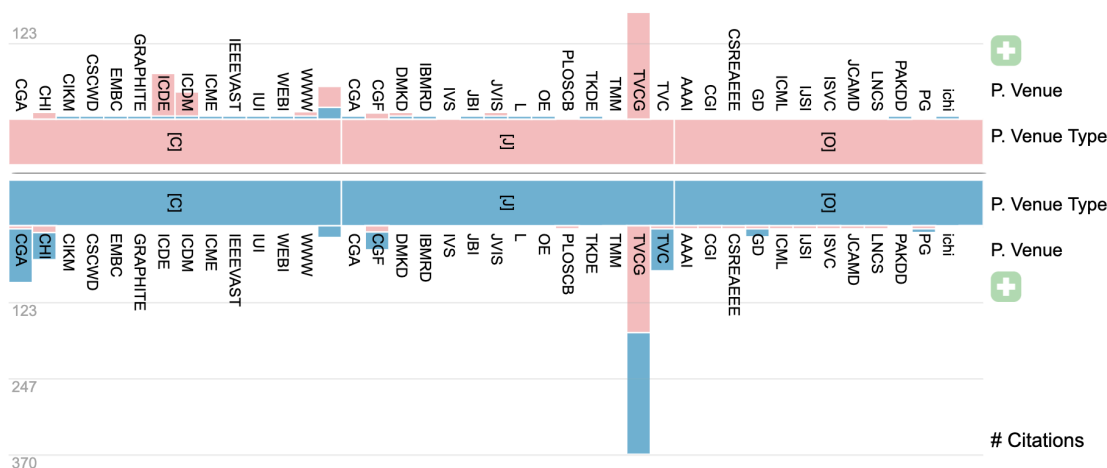

Figure 5. Aligning two hierarchical histograms for comparison.

- "**Bar alignment**": when bar alignment is enabled, the corresponding *horizontal bars at the coarsest level* in the upper and lower hierarchical histograms will be aligned. If no corresponding bar exists in the other histogram, an empty bar will be inserted. In addition, the shorter bar will be mapped to the taller one for easy visual comparison of the differences. For example, in Figure 5, the red "TVCG" bar in the upper histogram is also shown in the lower histogram to indicate its difference from the blue one. Otherwise, the bars will be aligned compactly for each histogram independently without any empty bar.
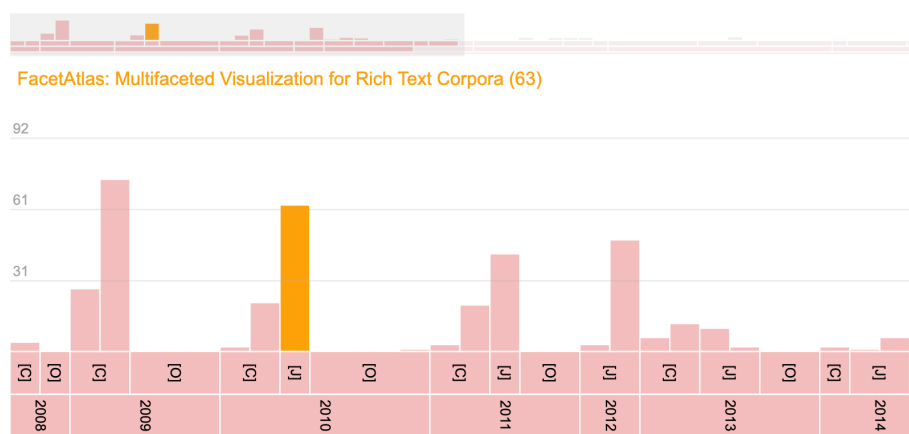

Figure 6. Panning and zooming a hierarchical histogram using a mini-map.

5

Panning and zooming with mini-maps. A **mini-map** is provided for panning and zooming the upper or lower histogram. As shown in Figure 6, by selecting the width of the viewing window, you may horizontally expand the histogram to view more details. You may scroll the histogram by dragging the light gray box in the mini-map. You may mouseover a bar to see its detail information, i.e., the name of the subset (in this example, the title of the paper as an individual paper is shown) and the height of the bar (in this example, the number of citations).
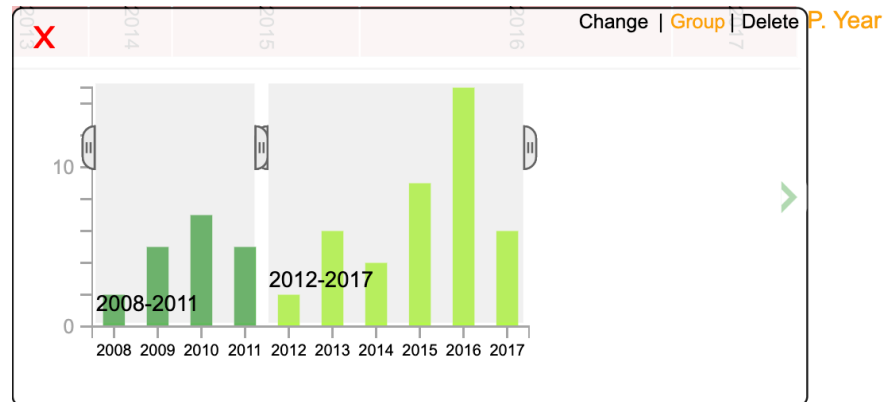

Figure 7. The interface to group multiple years.

Bar grouping. The bars can be further grouped to investigate the bars as a whole. For example, as shown in Figure 7, to group the publications years, you may click on "Group" in the pop-up menu when mouseover the "P. Year" attribute and brush the resulting histogram to form the groups. Once a bar group is formed, users can also remove it (acting as a filter). You can remove one group by naming it as "ignore". For example, users can update the "P. CCF Rank" attribute to form and remove bar groups, leaving only CCF rank A papers to be shown and explored in the hierarchical histogram.