Visualizing the Relationship Between Scholarly Entities

Vignesh Shankar, Deepika Teegapuram, Graduate Students, University of Regina

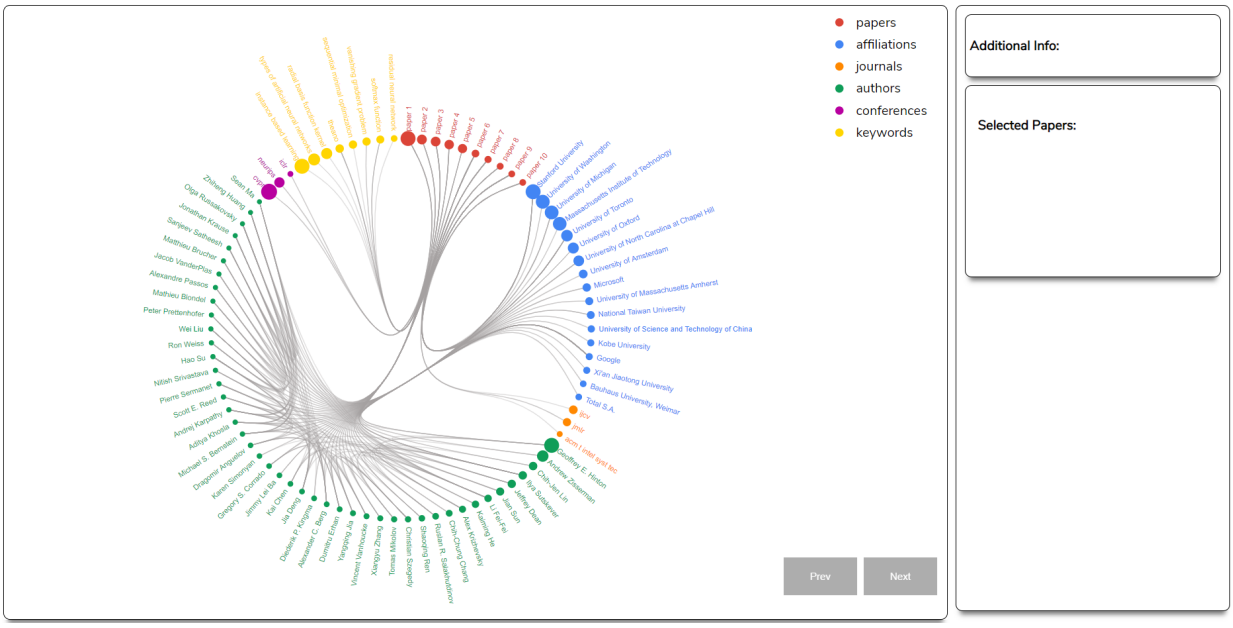

Fig. 1. Default view of the Relationship between scholarly entites.

Abstract—In todays world in each domain, the amount of data is increasing at an exponential rate. This effect is also applicable to the world of scholarly data. Due to no small amount of data, more researchers are focusing on the analysis of the domain and tend to lack in presenting the findings of the analysis qualitatively especially when it comes to representing multi-facet data for the applications based on comprehensive literature review. Hence we are going to implement a practical design for doing a literature review by addressing the issue of visual clutter and overlapping of visual variables.

Index Terms—Visual Clutter, Scholarly Data, literature review, effectiveness, Human perception

1 Introduction

Scholarly data is a heterogeneous hierarchical data that contains information about the published papers such as paper name, authors detail, a field of study, affiliation, venue, and event. This information is increasing day by day exponentially, which arises the opportunity for analyzing the scholarly data.

However, there are some issues with the data such as missing information due to publication copyright license and the absence of entity information such as conference names, journal names, and reference information.

The analysis of scholarly data is not only useful for academic research but also useful for commercial purposes. For instance, analyzing which institution or a country is focussing more on which field of study.

Although there is an enormous opportunity in analyzing the scholarly data, the visualization of this data is crucial as it provides the overview and comprehensive information of the scholarly data together as one view.

However, we can strongly argue that there is a lag in the visual representation of the overall structure and relationship between scholarly data entities. In particular, visualizing multiple scholarly facets leads to visual clutter and overlapping, which causes the data to hide and mislead the user.

Also, some of the applications/visualizations are having low data to ink ratio and representing more information than the user wants, which in turn makes these visualizations less effective and expressive.

For instance, the existing visualizations focus only on the high-level overview of a few entities of the scholarly data. There is no feasibility to add/customize new entities. Even if they do add, the visualization will turn more complex and difficult to understand, although the overall pattern can be judged but does not focus on the user's needs.

In this paper, we focus on effective design, which represents the relationship between all the facets of the scholarly data by following the principles of visualization foundations and taking into account of capabilities of human perception. Our main contribution in this paper is:

1. Creating a field of study centric visualization design which represents the relationship between scholarly entities interactively.
2. Organized the multiple facets in an appropriate manner such that they do not overlap with each other.
3. Combined text and visualization so that the user can gain more information about a specific facet of scholarly entity instances relationship in detail.

We are focussing on the design rather than the dynamical retrieval of information such that our primary objective puts more emphasis on

- *Vignesh Shankar, Student-ID:200415937 ,E-mail:vsh409@uregina.ca.*
- *Deepika Teegapuram, Student-ID:200415938 ,E-mail:dtu886@uregina.ca.*

creating a snapshot design for a default view of interconnecting all the scholarly data entities. However, it also focusses on the exploration of these entities as a secondary objective.

2 Related work

Various work has been done previously on visualizing the relationship between scholarly data entities such as PivotSlice[1], which implements an exploration of an implicit and explicit relationship between heterogeneous scholarly entities/facets by building dynamic queries. It also uses network visualizations such as a force-directed graph to establish the relationship between scholarly data entities. Moreover, they have provided an interactive visualization that allows the user to explore the information such as finding the Top venues for a particular field of study and also to find the trends. However, the application does not address the issue of visual clutter and low data-ink ratio. Also, before using this application, the user must be trained. Along the same lines, PivotPaths[2] implements a visualization application that represents a relationship between paper, authors, and keywords by an exploration of facet entities.

This application satisfies the ease of exploration, visual consistency, and shows the relationship between facets and focuses more on browsing more than searching. This procedure has been generalized for all multi-facet or multiple data such as movies, academic publishing, and YouTube video streaming clips. However, this application gets cumbersome when there are more facets of data. Due to which the problem arises such as overlapping of nodes which causes information hiding which confuses the user and reduces the effectiveness of the graph.

Several other visualization techniques to find a pattern or a trend in the data have been proposed such as citeRivers[3], that aims to find new trends in the citation pattern by linking the topics of the paper to the venue of the publication. It uses streamgraph r4 (also called a theme river model) Which links papers, topics, venues, and citations. There are five views of this visualization. In the first view, the stream graph panel uses NLP techniques to fetch the profoundly impacted words as the topics of the document and display it as a word cloud. In another view, only the top 10 authors of each document are displayed using a limited set of colors. At the end of the river, the relevant documents that are fetched based on the word cloud and the venues of each paper from the citation journey are shown. By looking simply at the visualization, one can note that different colors have been used to represent the authors who belong to the same domain, which leads to inconsistency. Overall, there are too many colors in this visualization. A better approach would be providing a legend which allows the user to segregate which one is a document and which one is an author. Similarly, VIS Author Profiles [4] has an interactive visualization that demonstrates the author's contribution throughout his career. This visualization has been divided into three sections.

1. General Information: This section discusses the author's published work in quantitative detail.

2. Research Areas: This section gives details about the field of study where the author has contributed his work and visualizing co-author work using temporal visualization.

3. Collaboration Relationships: This section gives information about the most collaborated co-authors and the number of research papers published with them.

This application also shows similar authors and their correlation of similarity. This visualization puts emphasis on the author's collaboration with other authors. Moreover, the author provides Interactivity combined with text and timeline visualization, which had more text compared to the visualization. If there were more visualization between the author and co-authors, it would have been intuitive and help the user to get an overview of rich insights. While we also have common objectives from previous work, but none of them focus on the design perspective problem such as visual clutter, node overlapping, expressiveness, and effectiveness of the design. Our objective is to build a visualization that addresses these issues by implementing a design that follows the principles of visualization foundations, Bertin's Semiology of graphs, and human perception.

3 Users Task and Goal

3.1 Problem Domain

In this project, we are focussing on the users such as students or researchers who have to go through numerous research papers to conduct a literature review on their research field before they start implementing a project. The user wanders around the available academic search engines and may end up selecting the papers which are irrelevant to their project. The Selection of a less impactful paper does not contribute to the research.

A better judgment on the paper selection can be made when the user understands more about its related scholarly entities such as the author of the paper, a paper associated with which affiliations, verifying if the paper belongs to a high-quality journal or conference and a papers subfield of study. This judgment can be inferred from establishing the relationship between scholarly entities.

Moreover, the user has to navigate each scholarly entity separately to get the related information which is frustrating and time-consuming.

In addition to that, the users find it hard to distinguish whether a paper belongs to journal or conference unless it is stated explicitly by the academic search engines. Although sometimes names of the journal or conference are self-explanatory, but there are some scenarios where we cannot differentiate for instance, ARIXIV:LEARNING this is a name of a venue for a research paper from this name of venue for a research paper, from this name we cannot judge itself if it is a conference or journal.

3.2 User's Goals

We have identified the users goal based on the issues mentioned above, which is:

Selecting relevant academic papers for conducting a comprehensive literature review in a field of study.

This goal will be further addressed in depth using the principles of interactive visualization in the below section.

3.3 User's Task

Below is the list of tasks that will help in accomplishing the desired goal.

1. While exploring, the user must be able to keep track of the papers that they have shortlisted for their literature review.

2. The user must be able to compare the quality of each entity element visually such that when the user hovers on to certain entity elements, the user will be able to compare the quality of its related elements and based on the quality of the elements, the user can gain some insights.

3. Users should be able to view more information about a particular entity and its related elements in an abstract manner.

4. In the default view, the user should be able to see a limited set of academic papers and must have the freedom to traverse forward for the new set of academic papers and backward for history.

4 Data and Scope

4.1 Data

To achieve the user's goal, we need to have a database that can retrieve information about all the scholarly entities. Since the scholarly data is evolving day by day, we need a database that grows with the growth of academic evolution. Instead of owning such a potential database, some online sources provide sufficient information about scholarly data. The following are some of the sources.

1. Google Scholar

2. Microsoft Academic Graph

3. IEEE explore

In this project, we have retrieved data from the Microsoft Academic Graph (MAG), which is a heterogeneous graph. It consists of documents published in multiple research domains and the relationship between those publications.

Initially, Microsoft launched this service as Microsoft Academic search and upgraded this version to Microsoft Academic Service and finally upgraded to MAG to handle the problem of missing information.MAG consists of six entities, namely:

1. Authors
2. Papers
3. Field of Study
4. Affiliation
5. Conference
6. Journals

MAG offers distributed service which releases a new version of the data whenever there is a change in the scholarly data due to the addition of the elements in the academic entities. We can access this distributed service either by connecting to Azure Storage Accounts or through rest APIs.

We approach the MAG's distributed service through the Rest APIs because they are lightweight processors. To fetch an entity and its associated entities, we need to use the Evaluate method of the Rest API.

To use this method for fetching any information related to an entity, we need to have an eye on the documentation provided by the Academic Knowledge API of MAG because we must know which attributes are having which information. E.g., To fetch the title of a paper, we need to use the Ti attribute of a paper entity. The Academic Knowledge API documentation provides information about the attributes associated with each entity. Also, it includes information about which other entity attributes it can accessed be from an entity.

The evaluate method takes three query parameters,they are:

1. Expr: Takes an attribute or attributes to filter the entities.
2. Attributes: these are the entity attributes which we want to fetch from a set of entities by relating with each other.
3. Count: the number of search results to be displayed in the response.

The evaluate method queries MAG data source by using the expr attribute and fetch the attributes (which are listed in the attributes parameter) of the entities by building relationships between them and displays the response number of records set in the count. For our visualization, we need papers and related entities of a field of study. Therefore, our expr parameter should be the field of research, and the parameters of the attribute should have the attributes of papers, authors, conference, affiliation, and journal entities. We built a database using Mysql to store the fetched results. The E-R diagram of the database is shown below.

4.2 Scope

Since there is a limitation on the evaluate method of the rest API, i.e., it can be requested only once per second. Therefore, we are restricting the field of study to be Machine Learning and fetching only 50 papers published in this field. For the user to see more relevant results, we are fetching the papers which are published from the year 2010 to 2019. As this is a snapshot design, we filter only the top twenty papers which have the highest citation counts from the fifty papers fetched.

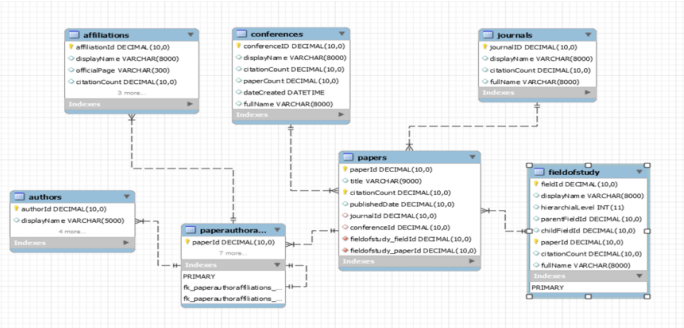

Fig. 2. The Entity Relationship Diagram of the database created

4.3 Transforming the data

The data related to 50 papers in the Machine Learning field of study is fetched from MAG and loaded into the Mysql database with the help of Python Programming. Using this database, we sort the papers based on citation count and only twenty papers. This data is transformed into a specific format of a JSON file using Python Programming such that this file can be used as an input for our Design Layout.

5 Design Layout for Visualizing Scholarly Entities

We have chosen a node-link d3-module diagram called Hierarchical edge bundling, which is used to visualize the hierarchical relationship between adjacent entities.

In this diagram, A node represents an entity element, and a link between one node and another node is represented by a curve or a straight line. This design will help us in visualizing multi-facet relationship with minimal visual clutter and overlapping.

This design layout focuses on the scope of the data, and the graph accommodates only ten papers in the Hierarchical edge bundling diagram to avoid visual clutter.

By default, this diagram shows the relationship between the entities for the top 10 highly cited papers retrieved from the database.

6 Design Requirements

Based on the user's goal set in the previous section, some design requirements should satisfy the goal. These requirements are listed as follows:

6.1 Distinguishability of Scholarly entities

To select the most impactful papers for literature review, we intend to use author, affiliations, sub-field of study conference, and journal as supporting parameters. All these supporting entities should be displayed in the visualization in such a way that at a glance, the user can understand the classification of entities.

6.2 Establishing visual ordering

The user can select a paper based on the standard or quality of the supporting entities, to show the quality of an entity we have to order the entity instance in the visualization. We can apply this ordering concept to all supporting entities of paper so that the overall visual consistency of the graph is maintained. For measuring the quality of each entity, there are numerous bibliometric measures available but considering the primary objective of our project, which is to build a practical design. We are considering citation counts as a quantitative measure. The Citation count is a common attribute for all the scholarly entities. Henceforth we are using citation count to show the order in the data in such a way that, each entity element is ordered in descending order of citation count within the scope of their entity and not relative to the citation count of the paper entity.

For example, all the authors in an Author entity are sorted by their citation count instead of the citation count of the associated paper, which makes the user understand where the author stands in comparison to other authors.

6.3 Relationship between Scholarly Entities

Once the data in each entity type is ordered in the graph, we should establish the visual relationships between the scholarly entities. The absence of this relationship will make the user opt for a manual search on the academic search engines to find the information of the paper, such as who are the authors of this paper. The paper is belonging to which venue and the paper associated with which affiliation and subfields of the paper.

In the visualization, the user should be able to see the collaboration between all scholarly entities for the paper to get the idea about how other scholarly entities associated with it. Using this approach, the user can see high-quality authors, conferences, and renowned institutions related to a paper. Based on this related information, a user can decide on discarding the paper. This design is useful to achieve the goal.

6.4 Interactivity

6.4.1 Highlighting

Providing the user to explore the data using Interactivity gives the user the freedom to understand the different entities' information abstractly and also interactively to highlight the related entities will make the user grasp the detail in a short period.

Highlighting the entity elements and their related entities will help the user to establish a relationship between scholarly entities. Meanwhile, visual consistency should be maintained in representing a specific entity with the same color such that the user does not have to always refer to the legend.

6.4.2 Information on Demand

In this design, there could be a situation of an ambiguous relationship between the entities. For instance, an institution has supported for publishing two papers. Consider that the authors of these two papers are different. As per our design, when that institution is selected, all the authors and papers from that institution will be highlighted. Here the user will not know which set of authors wrote which paper. Hence, the design should contain a section that clearly explains the information between paper and related author details in the text. This textual information should not decrease the importance of visualization.

Not all relationships between scholarly entities are ambiguous. Therefore, the extra information should be demanded by the user to see more detail to resolve the ambiguity. Overall, there should be a section that shows the textual information between entities on demand by the user.

6.4.3 Shortlist the papers

This design should give feasibility to the users to shortlist papers from the visualization. The shortlisted papers should be maintained in a separate section. Thus, it will make the users concentrate only on the shortlisted papers to avoid the search for the paper again and again. Also, When the user adds too many papers in this section, our design should support auto-scroll on this section. From this approach, we conclude that our design is divided into three sections that are, a diagram section which has the design layout, a section for textual information (Information on Demand) and finally a shortlist section.

6.4.4 Selection from shortlisted papers

When a user selects a paper from the shortlisted section, the details of the selected paper and its relationship with the associated entities should display in the Information on Demand section. It will allow the user to compare the quality and relevancy between the selected papers easily.

6.4.5 Deletion from Shortlisted Papers

Users should have control of deleting the shortlisted papers when they are not willing to use that paper. However, the user should always be able to add the deleted paper.

6.4.6 Navigation to the next set of papers

As the design layout accommodates only ten papers in the diagram, our design should not restrict the users to see only these ten papers, To fetch more relevant papers, the user needs to navigate for the next relevant papers such that its relationship with other entities are also loaded in the diagram in the design layout section, i.e., papers from 11-20.

This change in data should not affect the functionality of the information on-demand section and shortlist section. The content present in these sections should not disappear. The new data and the diagram should support all the interactive techniques. Similarly, the user should be given a chance to navigate back to view the previous paper's associated relationship with entities, i.e., and we reload the diagram with the papers from 1-10. Here we should be careful in addressing the following scenarios:

1. When a user shortlists a paper from the default view and then navigates to the next relevant papers, the added papers from the default view should appear on the second view in such a way that the user can select the paper of default view and still able to visualize its entity relationship in the additional info section. Similarly, the papers shortlisted from the other view should display on the default view, and when the paper is selected, it should describe its entity relationship in the on-demand section. In this way, consistency is maintained in the diagram.

2. Similarly, when the user deletes a paper irrespective of the view, the paper should be deleted from the shortlist section.

7 Visual Encoding and Interactive Mechanism

7.1 Design Layout

We have chosen a node-link d3-module diagram called Hierarchical edge bundling which is used to visualize the hierarchical relationship between adjacent entities. In this diagram A node represents an entity element and a link between one node and another node is represent by curve or a straight line. This design will help us in visualizing multi-facet relationship with minimal visual clutter and overlapping. This design layout focuses on the scope of the data.

7.2 Distinguishability of Scholarly entities

In scholarly data, there are six entities which are Paper, Author, Affiliation, Journal, Conference, and Keyword. We have used the color variable of Bertins semiology of association to distinguish between the scholarly entities. Hence, we have used six different colors to represent six different entities such that the user can perceive similar entity elements. We have selected the colors based on the opponent theory of color, which suggests that how our brain perceives color and can differentiate between three color channels which are Red-Green, Yellow-Blue, and luminance(black-white). Since HSB color space matches the principles of the opponent theory of color, Therefore, we are going to use HSB color space to select colors and also provided a legend which can help the user to interpret which color has been used to represent which enitity.

Moreover, We are using the gestalt principle of similarity to group all the elements of an entity together such that the user can easily identify an entity in the visualization with minimal effort.

7.3 Establishing visual ordering

Based on our requirements, we need to show ordering in the entity elements based on their citation count. The best way to do this is to show the difference of size on the shape of the circle by placing the circle adjacent to the entity element.

Larger the size of a circle of an element better its quality. Similarly, smaller the size of a circle of an element lesser the quality. Representing circles with the same color as their entity makes the visualization consistent, which further eases the effort of the user to compare the quality between the entity elements without much focus.

We are not showing brightness as a visual variable on a circle to show the ordering because the circles are already in association with

the entity by being in the same color Hence representing ordering by brightness will make the user pay more attention and focus than representing the ordering with size.

7.4 Relationship between Scholarly Entities

With the help of visualization, the user should be able to understand the relationship between the entities. We applied the gestalt principle of connectedness, which shows the relationship between two graphical objects by drawing a line between them. This principle is a powerful mechanism to show a user that relationship exists between two graphical objects.

In our project, we implement this principle by connecting the nodes which are related to each other with a curve. We are using curves instead of straight lines because curves provide more readability and enhance the visual appearance.

7.5 Interactivity

7.5.1 Highlighting

We are using Selection + Delete interactive mechanism to implement Interactivity in this visualization.

Using a mouse hover event on an element of an entity, we are highlighting the selected element along with its related entities and their connections which are represented by a curve.

The highlighting of the elements is done using the color visual variable of Bertin's semiology of Selection. For ease of understanding, we would be calling the selected element as a source element and all the other related elements as target elements. When the mouse has hovered on an element, the source element is highlighted in black color, and all the target element will be highlighted in their current color. Similarly, the curve which represents connections from source to target is highlighted in the color of the target entity.

Moreover, all the unrelated elements and their connections(curve) will disappear from the graph, and the opacity of their circle becomes translucent such that the data-ink ratio becomes high and the user will able to interpret relationship between source and target elements quickly with minimal effort.

Due to the presence of numerous elements in the graph, there is a chance that the entity element's name can get overlapped with other element's names while highlighting. To resolve the overlapping issue, we have kept the font size of element small, and also, we are displaying the paper name with a format such as papernumber to avoid displaying long title names. However, it arises a new problem that user will not be able to see anything with such font size. For resolving this issue, we have used a tooltip that displays the information of the source elements on mouse hover which are, the complete name of the element and its citation count.

Furthermore, on Mouseout, the default view which is all the elements and their connections with other elements are displayed again. Overall, by using the Selection + delete interactive mechanism, we can make our visualization expressive and effective and reduces visual clutter.

7.5.2 Information on Demand

This implementation is an extension of the mouse hover event. Although the mouse hover on an element of an entity highlights its related entities and their connections, sometimes this may not be enough to the user. For instance, the user is interested in knowing in which year the paper was published. This information is not shown in the visualization since the year is not an entity, but it is a metadata attribute of the paper entity. Therefore, we created an on-demand section in our visualization, which shows the metadata information of the selected element and its relationship with other entity elements in a textual format. To view this textual visualization, the user has to click on the element rather than merely hovering. In this way, the problem of metadata and ambiguous relationship, i.e., many-to-many relationship (discussed in the design requirement) is addressed.

Also, the mouse click option provides convenience to the user to read about an element, and at the same time, the user can explore other entities in the interactive visualization.

7.5.3 Shortlist the papers

Whenever an interesting paper is found through web-browsing, the user makes a note on it for future reference. In our design, we handle this by creating a section called a shortlist section that keeps track of the papers selected by the user. A right-click on the paper element adds that paper to this shortlist section. This section is placed just below the Information-Demand section. There are multiple validations applied in this section such that when a paper is already present, a paper with the same name should not enter inside. This section is built on the gestalt principle of closure, such that all the papers are present inside a boundary.

7.5.4 Selection from the shortlisted papers

Within the shortlisted section, when a user left-clicks on a paper, the relationship of all the entities with the selected paper should be displayed in the Information-on-demand section. In addition to that, when a user hovers on a paper in this section, the hovered paper will be highlighted based on the Mackinlay's visual attribute of containment.

7.5.5 Deletion from the shortlisted papers

Similar to the above functionality, when a user realizes that the shortlisted paper is no longer useful for their work, they may want to delete it from the shortlist section. We have implemented this functionality such that whenever a user right-clicks on the paper in this section, that paper will be deleted only from this section, not from the visualization.

However, the user can again add that paper to the shortlisted section by performing left-click on the paper element in the visual diagram.

7.5.6 Navigation to the next set of papers

As per the requirement specified in the design, we created two buttons called prev and next for traversing back and forth for fetching the more papers.

When the prev button is clicked by the user on the default view, it throws an alert message because it is the first set of papers. When the next button is clicked on it, papers from 11-20 are displayed. On the second view, when the prev button has selected, the papers ranging from 1-10 are displayed. On any view, the user can add papers to the selected papers section and can select the paper by left-clicking on it to see the information on additional information section.

Similarly, in any view, the papers in the selected papers section can be deleted by right-clicking on the paper.

8 Results

For evaluating the results, we are creating a hypothesized scenario where the user is a novice student or researcher. The user needs to select relevant academic papers and conduct a literature review in the Machine Learning Field of Study.

We are going to evaluate the above scenario based on our design requirements.

8.1 Distinguishability of scholarly entities

We can see from fig[3] that six entities can quickly be perceived from their color. All the elements present in that entity are grouped. Also, the color used is easily distinguishable from one entity to another. From all this, we can say that this design has achieved the goal of this requirement.

8.2 Establishing Visual Ordering

From fig[3], we can see that the user can interpret each entity element is in descending order on the size of the circle and contains some color as their entity type. The tooltip provided on hovering the element, states the citation count of each element. Therefore, we can understand that visual ordering is on the citation count. Also, the circles are neither overlapping with any other circle nor overlapping with the text of the entity elements.

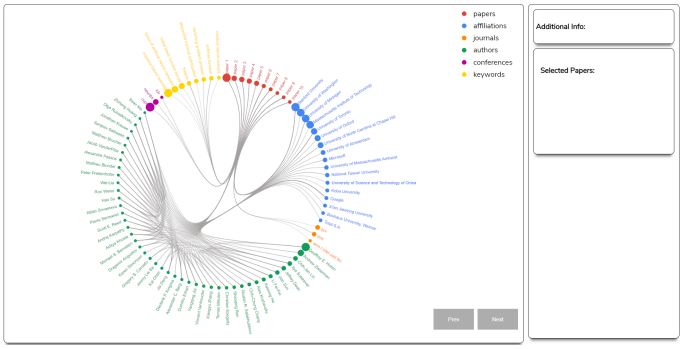

Fig. 3. The Entity Relationship Diagram of the database created

8.3 Relationship between scholarly entities

From fig[3], the user can see from the default view that the relationship between the entities using curves. However, the relationship of an individual entity element can be viewed more clearly with the help of interaction.

8.4 Interactivity

8.4.1 Highlighting

From fig[4]], we can see that the relationship of all the scholarly entities with the paper. Along with this, when a mouse has hovered on an element, the hovered element's text and its related element's text is highlighted, and the curve connecting to them are displayed based on the target entity's color. Along with that, a tool-tip with the element name, citation count, and message for an appropriate mouse click event is displayed. The unrelated elements and the curves associated with them are disappeared, and the circles belonging to the unrelated elements decreased their opacity. Furthermore, on mouse out, the default view is displayed. Finally, the relationship provided in fig[4] is satisfied.

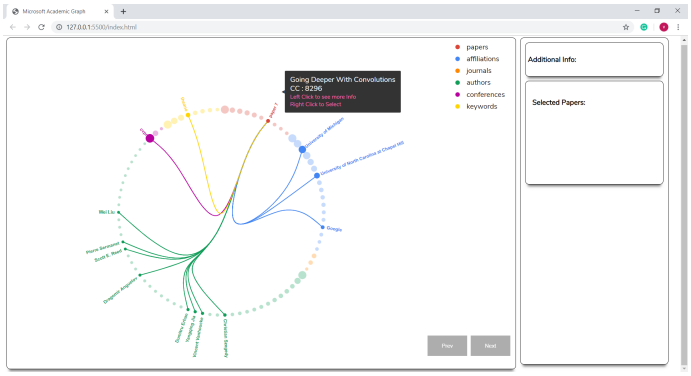

Fig. 4. Highlighted on a paper

8.4.2 Information on-demand

The user some times may not understand which author wrote which paper. After left-click on that element fig[5], shows the clear textual representation of the relationship of that entity element.

8.4.3 Shortlist the papers

The right-click on the paper elements on the first and default view are stored in the shortlist container, from fig[6] we can see that a paper has been shortlisted successfully.

8.4.4 Selection from the shortlisted papers

Irrespective of which view it is, when mouse left-click is performed, the related content of that paper is displayed on the additional-info section.

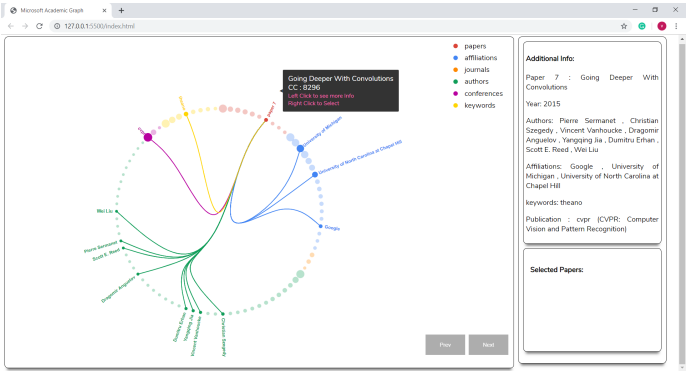

Fig. 5. Additional information about the paper is given

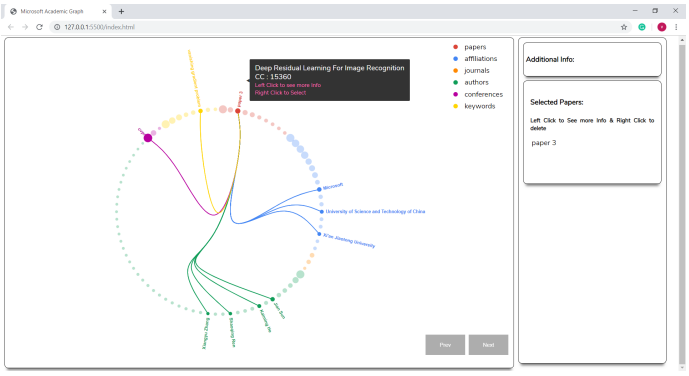

Fig. 6. A paper has been shortlisted in a selected paper section

in fig[7] we can see that the paper from shortlisted section is selected and the related information is displayed on the additional info section.

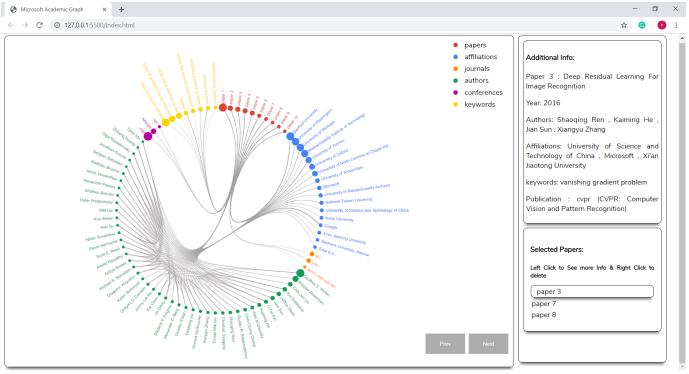

Fig. 7. The information of the selected paper is displayed in the additional info section.

8.4.5 Deletion from the shortlisted papers

Irrespective of which view it is, when mouse right-click is performed on the shortlisted section, it deletes that corresponding paper. The delete has been done successfully.

8.4.6 Navigation to the next set of papers

when clicked on next, papers ranging from 11-20 are displayed and when prev has clicked the papers ranging from 1-10 are displayed. All the interactive techniques are working correctly. fig[8] shows the navigation to the next set of papers, fig[9] shows the selected paper of default page in the second view. Similarly fig[10], the paper from second view is shown in the default view.

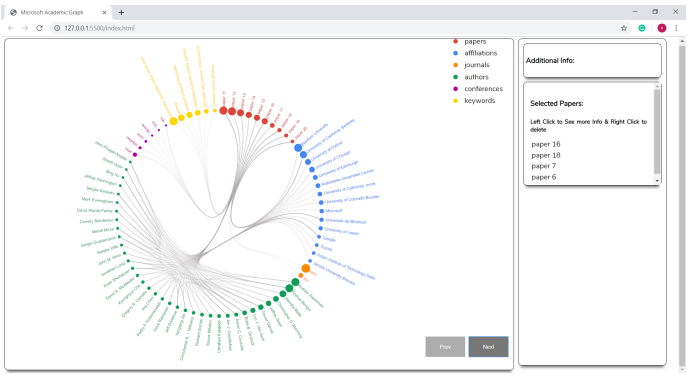

Fig. 8. Navigation to the next set of paper

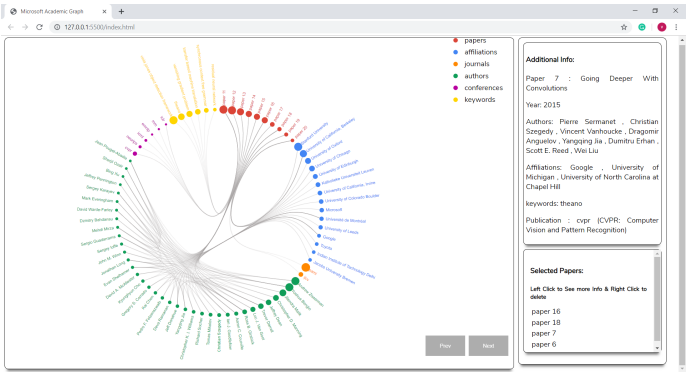

Fig. 9. The selected paper of default page in the second view.

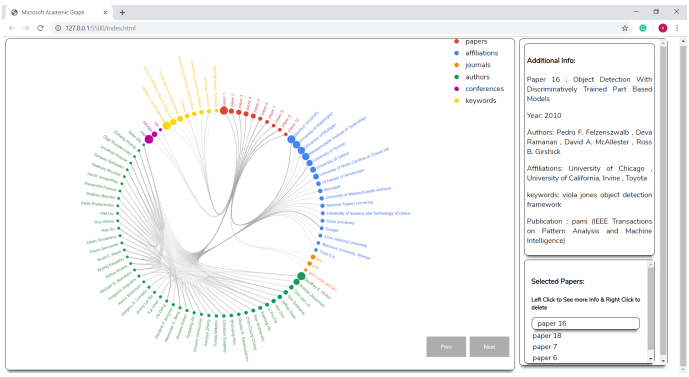

Fig. 10. The paper from second view is shown in the default view.

9 Benefits and Limitations

9.1 Benefits

Benefits Our design can reduce user efforts in selecting a relevant academic paper by using the principles of visualization foundations and using the human perception method. Our application shows all the information and only the information required for the user, which in turn gives us a steady hand in reducing visual clutter where most of the other researchers tend to lack.

In addition to that, our design can represent six categorical variables and their related elements without overlapping. In our application, we have given the user the freedom to explore the data such that the user can test their hypotheses and find an interesting feature. Also, keeping in mind the capabilities of the human memory system, we have given a User Interface panel where the user can add the academic paper which they might think relevant. Thus, our design is providing the user to explore the data with minimal effort.

9.2 Limitations

Limitation We can show order using the size of a circle, which in turn tells the user the quality of an entity. However, it is an effective method, up to an extent only the user can decode the differences in size. Which is why we have ordered the elements in descending order by their citation count in the Hierarchical edge bundling.

As to design a visual application for conducting a literature review, we have chosen citation count as the variable to measure the entity quality. However, there are more effective methods such as Eigenfactor,h-index,i10, and each entity has its measuring variables (all calculated based on citation count).

10 Future Work

Further studies on this area should focus and investigate how to visualize forward and backward cited papers on top of this visualization, which can further reduce the users effort. In addition to that, filter-based search engines can be used so that the user can search other papers based on other scholarly entities, which again gives more freedom to the user to explore.

11 Conclusion

In summary, our goal was to design a visualization application where the user can select relevant papers to conduct a comprehensive literature review. As the information is increasing day by day on this domain, visualization of these data, or to establish a relationship between them becomes more complicated. Due to the complexity of the data, most of the visualization ends up in overlapping of visual attributes which in-turn makes the visualization a victim of visual clutter. So we implemented a design that addresses these issues and accomplishes our goal. Our design can represent relationship six entity effectively without overlapping of the visual attributes which were used to represent the entity and also it allows the user to explore the data interactively with minimal focus and attention. In addition to that, we have used interactivity between text and visualization, which provides the user with rich insights. As we are using a small subset of data to represent this visualization, this design can further be improvised by visualizing the forward and backward citations and also some text analysis can be done to add more insights.

References

[1] Zhao, J., Collins, C., Chevalier, F., Balakrishnan, R. (2013). Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets. IEEE Transactions on Visualization and Computer Graphics, 19(12), 20802089.

[2] M. Dork, N. Henry Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through faceted information spaces. IEEE Transactions on Visualization and Computer Graphics, 18(12):27092718, 2012. doi: 10.1109/TVCG.2012.252

[3] Heimerl, F., Han, Q., Koch, S., Ertl, T. (2016). CiteRivers: Visual Analytics of Citation Patterns. IEEE Transactions on Visualization and Computer Graphics, 22(1), 190199.

[4] Latif, S., Beck, F. (2019). VIS Author Profiles: Interactive Descriptions of Publication Records Combining Text and Visualization. IEEE Transactions on Visualization and Computer Graphics, 25(1), 152161.