Visualizing Feature-based Similarity for Research Paper Recommendation

Corinna Breitinger
Dept. of Computer and Information Science, Univ. of Konstanz & University of Wuppertal, Germany
breitinger@uni-wuppertal.de

Harald Reiterer
Dept. of Computer and Information Science
University of Konstanz, Germany
harald.reiterer@uni-konstanz.de

***Abstract*—Research paper recommender systems are widely used by academics to discover and explore the most relevant publications on a topic. While existing recommendation interfaces present researchers with a ranked list of publications based on a global relevance score, they fail to visualize the full range of non-textual features uniquely present in academic publications: citations, figures, charts, or images, and mathematical formulae or expressions. Especially for STEM literature, examining such non-textual features efficiently can provide utility to researchers interested in answering specialized research questions or information needs. If research paper search and recommender systems are to consider the similarity of such features as one facet of a content-based similarity assessment for academic literature, new methods for visualizing these non-textual features are needed. In this paper, we review the state-of-the-art in visualizing feature-based similarity in documents. We subsequently propose a set of user-customizable visualization approaches tailored to STEM literature and the research paper recommendation context. Results from a study with 10 expert users show that the interactive visualization interface we propose for the exploration of non-textual features in publications can effectively address specialized information retrieval tasks, which cannot be addressed by existing research paper search or recommendation interfaces.**

Keywords—Recommender systems, interactive information retrieval, feature analysis, information visualization, user studies

I. Introduction

Scientists and academics rely on search and recommendation systems to discover and make sense of the most relevant or the most recent publications in their field. Prominent examples of academic search engines are Google Scholar and Semantic Scholar, which, in addition to search also display recommendation lists for registered users. Today, nearly all digital libraries, such as the ACM Digital Library, IEEE Xplore or PubMed Central (PMC), offer built-in recommendations to support the discovery of related articles while browsing (cf. Fig. 1). For the research paper recommendation scenario, the most common results visualization technique remains the ranked list [1]–[3]. Ranked lists condense the basic metadata of potentially relevant research papers for readers, such as the publication's title, author names, and sometimes the publishing date or a preview of the abstract. List-based visualizations of results are intuitive and simple to implement. For example, they can even be implemented with the help of external service providers for generating recommendations [4].

However, any fine-grained similarity present in the content of documents must be explored *manually* and with great effort by the researcher, since existing paper recommendation system interfaces rarely go beyond a ranked list visualization. Thus, more granular features of similarity, such as academic citation-based similarity, figure-based similarity, image-based similarity, or mathematical formula-based similarity remain undiscoverable using today's literature recommendation systems.

Note that for the paper recommendation use case, we equate high *similarity* of a paper's content features with a higher potential *relevance* to the reader. The reason for this is that considering the feature-based similarity among papers allows researchers to answer specialized information needs, such as: (a) discovering articles that contain identical or similar mathematical formulas to an examined article in question, (b) discovering articles that contain similar figures or charts, or (c) determining the fraction of shared academic citations among one or multiple articles, as well as their citation contexts within the articles being recommended.

The inadequacy of visualization approaches in existing research paper recommendation systems is further substantiated by the way in which researchers interact with recommended literature. In contrast to movie, music, or product recommendations, academics are not always interested in 'consuming' in its entirety the content of a multi-page research paper. Rather, researchers may be interested to grasp the condensed results, discover the datasets used, or find a particular insight or fact to answer a research question. In a 2018 survey by Tenopir et al., only around one third of respondents from the *computer science*, *engineering* and *science* domains stated that they read *all* of their most-recently-read research article 'with great care' [5]. In this study, the majority of STEM-domain researchers stated that they (1) read only parts with great care, that they were (2) only looking for the main points, or looking to find (3) specific sections or figures. Interestingly, respondents from the STEM fields were more likely to state that they were only interested in reading and understanding specific sections or figures, with 11.8% of the science disciplines and 3.9% of computer science disciplines choosing this option (out of 5 choices). In contrast, only 1% of respondents for the *social sciences* selected this option as a primary aim when reading the literature [[5] p. 9].

We summarize that the utility for academics does not necessarily arise from the consumption of the entire recommended document, but rather occurs upon discovering the content that helps a researcher answer a more specialized information need. Thus, a more inclusive and user-driven visualization of the content present in a set of recommended publications is crucial. Additionally, we conclude that the desire to quickly identify certain sections or figures within the literature is especially pronounced in the STEM fields, as indicated by the survey of Tenopir et al. [5]. To help researchers discover all features contained in a publication more effectively, we examine how researchers may benefit from novel visualization approaches, which in addition to traditional textual similarity, also visualize the feature-based similarity that can be of interest to a researcher. Specialized visualization approaches in academic literature recommendation can help guide the researcher to the specific features that matter most to them and their information needs. We define the document features unique to academic publications to include *citations*, *mathematical formulae*, *figures*, *charts*, or *tables*, in addition to *textual*, i.e. keyword-based similarity. In the academic recommender domain, the visualization of these features has thus far not been given much attention. Our research contribution is twofold:

(1) First, we review existing systems capable of visualizing instances of *feature-based similarity* among documents. We find that visualizations for document features have hardly been applied to the *research paper recommendation* use case. Due to the dearth of research, we additionally examine visualizations that have been proposed in related contexts, such as the near-duplicate detection and plagiarism detection (PD) domains.

(2) Second, we present a user-customizable *visualization interface* capable of exploring and explaining the above-mentioned features contained in sets of recommended research papers. In an expert evaluation, we examine how the proposed visualizations can support users in making sense of these unique features.

The remainder of the paper is structured as follows. We first review existing visualization techniques that have focused on the visualization of document features beyond textual similarity alone. We subsequently introduce our visualization concepts for feature-based similarity, which we have implemented in RecVis [3], our research paper recommendation prototype. We present the results of an expert study and lastly discuss plans for future research.

II. Background

In this section, we review two research areas with relevance to our work: visual approaches for recommender systems and visualization techniques that support the discovery of similar semantics features, i.e., techniques from the near-duplicate detection or PD domain.

A. Visualization Approches for Recommender Systems

The visual interfaces of nearly all research paper recommender systems make use of ranked lists to display recommendations to their users [3]. Figure 1 shows this standard view for two common libraries: the ACM Digital Library (left) and IEEE Xplore (right).

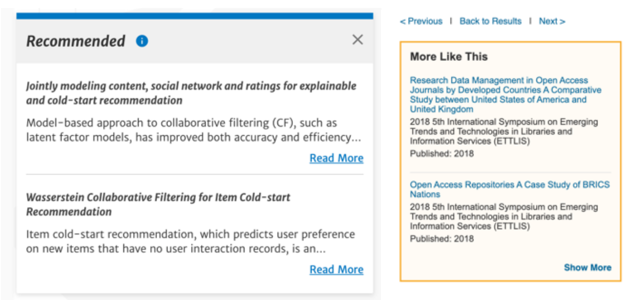

Figure 1. Standard visual interface for research paper recommendation.

This traditional recommendation interface displays a list of items ranked by predicted relevance and reduces the cognitive overload associated with exploring a more rich set of items [6]. However, list-based views come with a range of weaknesses, such as limiting the number of results using an arbitrary cut-off value, e.g., top-2 or top-5. Furthermore, there is no transparency as to which criteria or forms of similarity are being considered. Lacking transparency can lead to lower levels of acceptance or trust in the recommendations [7]–[9]. Lastly, list-based formats are ill-suited for an interactive exploration of specific features of relevance, especially because no existing interfaces provide a more high-resolution visualization of the recommendation sets to support a quick identification of specific sections or non-textual features.

The limitations of the ranked list and of existing paper exploration interfaces motivated us to examine approaches for visualizing document relevance using feature similarity, to provide readers with a more fine-grained resolution. Research on recommender systems has traditionally focused on improving recommendation accuracy. Only more recently have researchers examined the importance of *explanations* and *visualizations* for increasing the transparency of recommendation results [9], [10]. Recommendation systems that ignore results visualization and explanations risk being perceived as an opaque 'black box' by users [11]. In addition to transparency, a second design characteristic influencing user experience is *interactivity* [12]. Interactive recommendation interfaces enable a user-driven exploration of the results, e.g., allowing for visualization parameters or personalization settings to be changed.

In the research paper recommendation domain, *Scienstein* was the first interface to support interactivity and transparency through user-customizable filters and settings [13]. Users of the system could adjust the hybrid recommendation system to consider a collaborative filtering approach, a citation-based approach, or a content-based similarity approach for paper recommendation. The visual recommendation interface offered a clustered icon layout of recommended papers (see Fig. 2) in addition to a ranked list view. Users could also further narrow the results by publication year and impact factor.

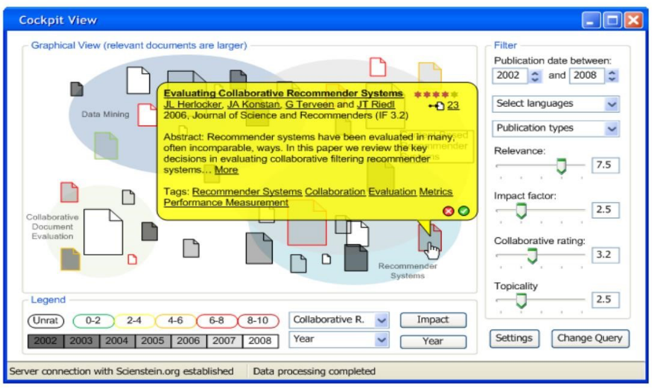

Figure 2. Scienstein [13] research paper recommendation interface.

While the Scienstein interface from 2009 was the first to propose that researchers filter the recommended papers according to their personal preferences, the system provided no supportive visualizations to display any instances of topical similarity or feature similarity within the recommended literature set. Users could not compare papers side-by-side or browse the features that may have contributed to their high relevance score and thus their recommendation.

The visualization tool *Connected Papers* [14] lets researchers upload a research paper of interest and builds a graph of similar papers using the Semantic Scholar Paper Corpus[1]. Each paper is visualized as a node and arranged according to its similarity score in a force directed graph (see Fig. 3). The authors use Co-citation and Bibliographic Coupling in their similarity metric, so even papers that do not directly cite each other can be connected and closely positioned. To increase explainability, the exploration interface of Connected Papers uses different shades of green to visualize the publication year and the graph-based layout arranges publications closer or further away depending on their similarity score. All recommended publications in Connected Papers are listed below the original input paper in a left-hand panel. By clicking on a list item or node, its abstract is displayed in a panel on the right, cf. Fig. 3. This visualization allows researchers to quickly examine the metadata and abstract. However, similarly to *Scienstein*, no features of interest in the suggested documents are visualized to researchers for further exploration.

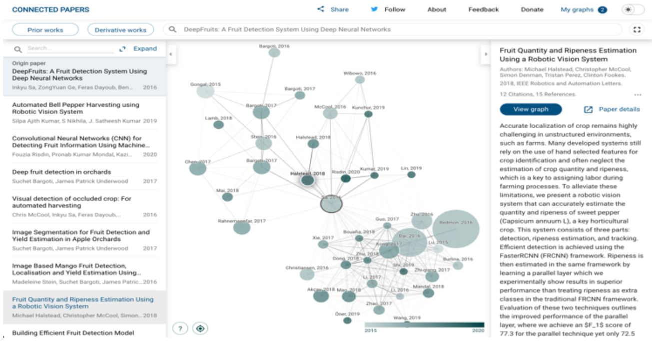

Figure 3. Connected Papers [14] graph-based visualization.

Parra et al. introduced *SetFusion*, a visual user-controllable interface designed for hybrid recommender systems combining Venn diagrams and sliders [15]. The system allows customized weighting of the different recommendation methods and visualizes the source of the recommendations using Venn diagrams, see Fig. 4.

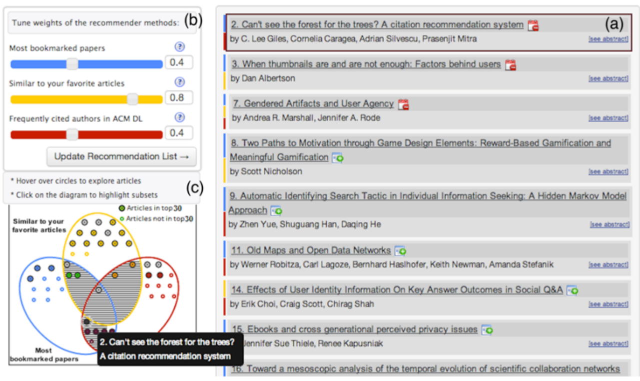

Figure 4. SetFusion [15] recommender with user-customizable weightings and a Venn diagram visualization.

This visualization approach shares the same shortcoming as the previously mentioned visualization interfaces. It allows for an overview of recommended items, but it does not support a detailed comparison of the item's *features* within the recommended documents in order to explore specific sections of interest or quickly identify the forms of similarity present.

Since the proposals for interactive recommendation interfaces to support the paper recommendation use case are sparse, we now briefly shift our attention to visualization approaches that have been proposed in *other recommendation contexts*.

In the *music* recommendation context, Bostandjiew et al. introduced *Tasteweights* [16], an interactive hybrid recommendation system to discover bands using social and semantic web resources, such as Wikipedia, Twitter, and Facebook. With their hybrid recommendation approach, they introduced an interactive interface see Fig. 5) that allows users to add or remove context and use sliders to filter the results. An expert evaluation of their interface showed that user-driven interaction with a visual representation of the hybrid system increased user satisfaction with the predicted content.

Figure 5. TasteWeights [16] user-customizable ranked list visualization.

In the *research talk* recommendation context, Verbert et al. introduced an interactive cluster map visualization approach in *TalkExplorer* [6]. Their proposed interface visually combines multiple 'relevance prospects', i.e., potential dimensions of relevance, by including bookmarks of users, suggestions by recommender agents, or talks marked by specific tags. The user can visually explore and combine these three relevance prospects by selecting or deselecting them, which adjusts the central cluster map visualization, see Fig. 6. Results acquired via questionnaires indicated that users perceived the visualization of TalkExplorer as valuable because they could gain more insights into *why* talks were recommended in comparison to a list-based format for recommended items.

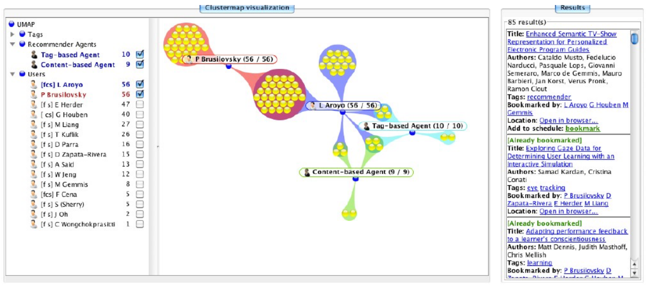

Figure 6. TalkExplorer [6] cluster map visualization for discovering research talks.

Other interfaces that proposed a richer exploration of recommendations include *PeerChooser* [17] for the movie recommendation context, *SFViz* [18] for the friend recommendation context, and *SmallWorlds* [19] for the social recommendation context. However, no existing interfaces support a *detailed comparison view*, i.e., in the form of a side-by-side exploration of the features contained in the sets of recommended items. This is all the more surprising when we consider that in the PD and near-duplicate detection

[1] https://www.connectedpapers.com/

domains detailed exploration and comparison interfaces are widespread to help users make sense of the recommended documents [20], [21].

B. *Visual Approaches for Document Similarity Detection*

In this section, we review sophisticated visualization techniques that have been introduced in the document similarity analysis domain, specifically plagiarism analysis for examining feature similarity within documents. While the discovery and visual exploration of the unique feature-based similarity within documents is currently not supported by recommender systems, several advancements in the visualization of feature-based similarity in academic documents have been proposed in the plagiarism analysis domain.

Meuschke et al. [22] and Gipp [23] describe the *CitePlag* prototype, which became the first examination interface to enable a visualization of citation patterns within an article's full text. Fig. 7 shows the introduced citation pattern visualization panel (center) where matching citations are highlighted in the same colors within the text to support users in inspecting potential instances of plagiarism that go beyond textual similarity alone. The premise underlying CitePlag is that similar or identical citation patterns can point to potential cases of idea plagiarism, translated plagiarism, and other disguised forms of academic plagiarism – even in the absence of verbatim textual overlap.

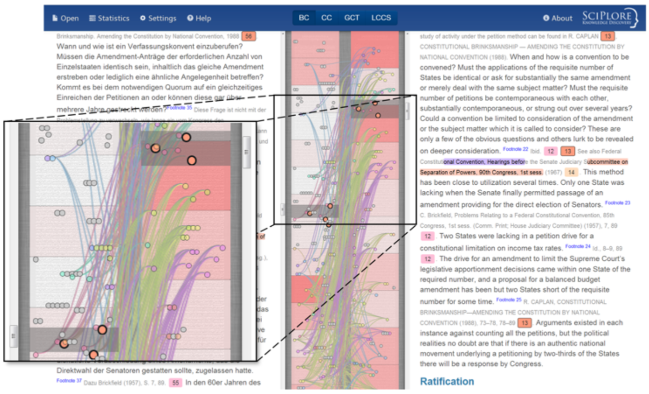

Figure 7. CitePlag [23] academic citation visualization to support the inspection of potential plagiarism.

By clicking on the circles representing matching citations in the central 'document browser' panel, users can jump interactively to the sections in the respective full texts of both documents where the citation match is found. In addition to the visualization of matching citations and their patterns, CitePlag also allows for the inspection of textual similarity in the same way as standard plagiarism inspection systems by highlighting matching keywords and text fragments. While CitePlag's visual interface allows for a detailed inspection of citation matches, if such a visualization were to benefit the literature recommendation scenario, the interface's capabilities must be refined to support a broader range of potentially relevant semantic features that academics may want to discover and explore.

Riehmann et al. [24] introduced an interactive visual analysis tool using a networked flow diagram to support the inspection of high textual similarity indicative of potential plagiarism. In Fig. 8, the left-most vertical line lists the potential sources that may have been plagiarized by the document being inspected. The central vertical line corresponds to the page numbers in the inspected document. The line-density for a certain page in the Sankey diagram thus visualizes the number of sources potentially plagiarized from on that page. The tool by Riehmann et al. also provides a detailed line-by-line inspection to examine instances of high similarity side-by-side (shown below the Sankey diagram in Fig. 8). Finally, the visual interface provides a scale for different plagiarism types, e.g., obfuscation, translation, or shake and paste, and different plagiarism types are color-coded for quick differentiation.

The graphical visualization by Riehmann et al. is restricted to examining text fragments, thus only visualizing structural and textual similarity. While such a visualization concept could allow for a detailed inspection of matching keywords among a set of recommended results, users cannot examine non-textual features. Furthermore, the comparison view is designed for an expert user with experience in the examination of plagiarism cases. The complex similarity inspection interface is not suitable for a reader or researcher interested in exploring the features contained in research papers to address a specialized information need.

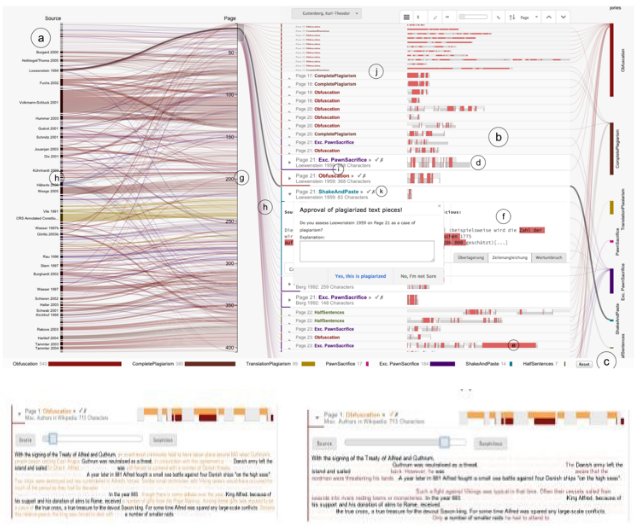

Figure 8. Sankey diagram used by Riehmann et al. [24] to visualize textual similarity in the PD context.

We conclude our review of related work with a plagiarism analysis and examination system introduced by Meuschke et al. [20]. *HyPlag* is the first system to extend similarity assessment and visualization to the full range of features uniquely found in academic articles. The HyPlag interface lets users upload documents for inspection and subsequently generates a 'results overview' and a 'detailed inspection' view. The backend implements feature-based assessment approaches for: mathematical similarity, image similarity and citation similarity, in addition to textual similarity computation. Users can interactively choose the similarity algorithms used and their sensitivity thresholds to more effectively filter among the different forms of similarity in the detailed inspection view.

More importantly, the inclusion of non-textual features for similarity assessment in the plagiarism detection scenario allows identifying even heavily disguised forms of plagiarism, such as translated plagiarism, paraphrased plagiarism, or idea plagiarism under the assumption that non-textual features, such as figures, charts, formulas, and academic citations are less prone to being diligently replaced by plagiarists.

Fig. 9 shows the results overview and the subsequent detailed plagiarism inspection view in HyPlag. The visualization concept of HyPlag is among the most comprehensive for the plagiarism inspection use case, since it supports multiple feature-based similarity analysis methods and supports the visualization of the full range of features that can be indicative of plagiarism in academic documents.

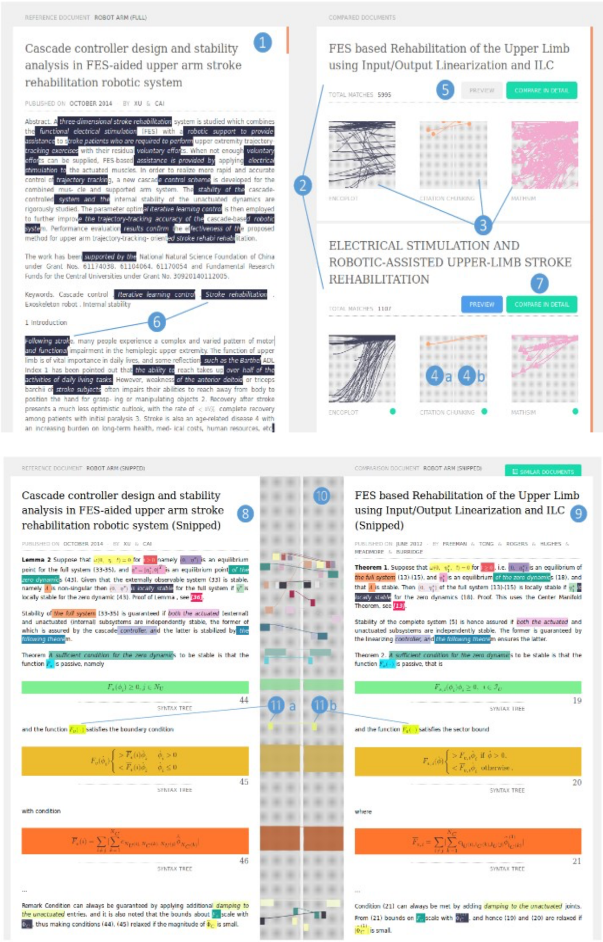

Figure 9. HyPlag [20] results overview for papers (top) and detailed inspection view (bottom).

Since the document feature visualizations in HyPlag are designed for plagiarism inspection, the interface is not optimized for the recommendation exploration use case where readers are interested in topical relevance or semantic equivalence rather than exact matches. Thus far, no feature visualization concept has been proposed or evaluated for the paper recommendation use case.

Table 1 summarizes the capabilities and weaknesses of existing exploratory interfaces we described in this section. We see that efforts have been made to address explainability and interactivity in the paper recommendation scenario. We also see that feature comparison interfaces exist in the plagiarism detection scenario. However, there are currently no interfaces combining these functionalities and applying them to the research paper recommendation scenario. We conclude that detailed feature-based visualizations for the *research paper recommendation* domain remain a research gap.

III. Visual Exploration of Research Paper Recommendation Sets

Having identified the shortcomings of existing visualization techniques, we dedicate the remainder of this paper to our second contribution. We conceive and evaluate a set of supportive visualizations for each of the previously identified aspects of feature-based similarity in STEM literature, i.e. citations, textual keywords, figures, and mathematical expressions. We implement these customized feature-based visualizations in the RecVis paper recommendation frontend, which we previously introduced in [3].

To provide the necessary context, we briefly review the graph-based visualization for exploring sets of recommended research papers, which we introduced in the RecVis prototype. Figure 10 shows the high-level overview for interactively exploring recommended papers and filtering the recommendations according to their contained features. The recommended papers are arranged as nodes of a force directed graph according to their relevance scores, where more closely related papers are visualized using shorter and thicker edges than more distantly related papers. Papers are recommended with reference to the user-selected central node that represents a paper, or a set of papers representative of the user profile, for which the user wishes to receive recommendations. It is this overview of recommended papers that precedes the detailed comparison of semantic features – which we present for the first time in the remainder of this section.

TABLE I. Overview of Visualizations Supporting an Exploration of Individual Features

Recommender System	Domain	*Explanatory interface*	*User customizable*	*Visualization of feature similarity*	*Detailed feature comparison*
Scienstein [13]	Research papers	✓	✓	✗	✗
Connected Papers [14]		✓	✓	✗	✗
SetFusion [15]		✓	✓	✗	✗
TasteWeights [16]	Music	✓	✓	✗	✗
Talk Explorer [6]	Research talks	✓	✓	✗	✗
Proposed System	*Research papers*	✓	✓	✓	✓
Similarity Detection					
CitePlag [22]	Similar document detection / PD analysis	✓	✓	✗ [a]	✗ [a]
Riehmann's prototype [24]		✓	✓	✗ [b]	✗
HyPlag [20]		✓	✓	✓	✓

[a] text only [b] citations and text only

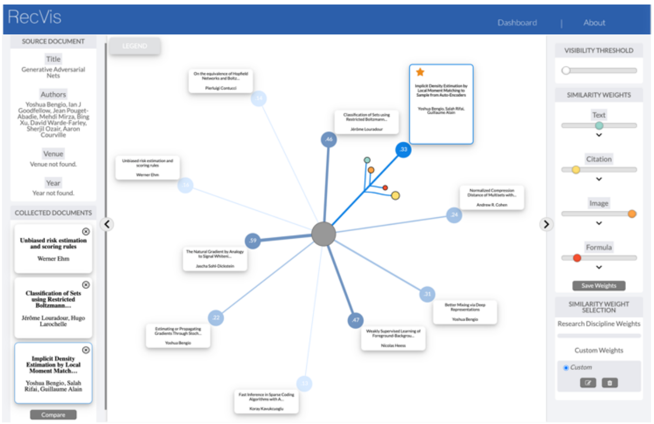

Figure 10. RecVis uses a graph-based layout to visualize recommendation sets.

Figure 11 shows a close-up of the interactive filtering of recommended papers in RecVis according to the similarity of their contained features. Once the recommendation set has been adjusted according to the user's information seeking need, the user can 'collect' the papers they wish to subsequently compare in more detail by starring them. From these user-curated 'Collected Documents' (cf. bottom of the left panel in Fig. 10) the user clicks a 'Compare' button to open the detailed feature comparison views, which we describe in the following sub-sections.

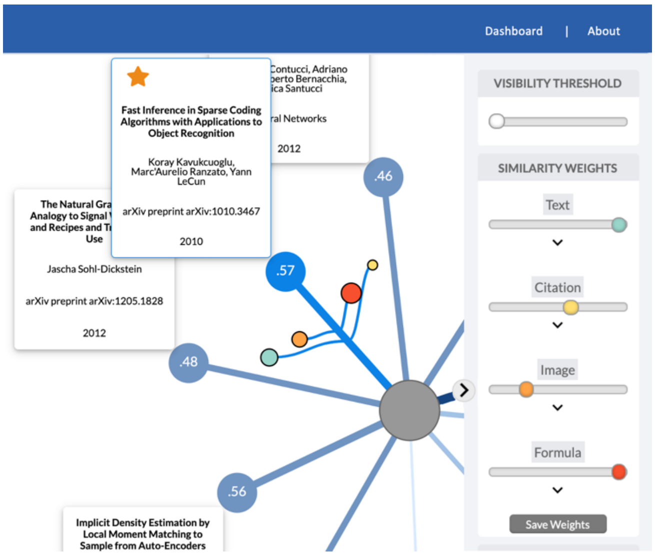

Figure 11. Recommendation sets are filtered by users according to their contained features.

Due to space limitations and the focus of this paper on the visual interface, please refer to [3] for the description of the RecVis implementation. In the following, we present the detailed feature-based visualizations to support the paper recommendation use case, which have not been proposed or evaluated thus far. Section 4 examines the perceived usability and effectiveness of the feature-based paper comparison interface in a study with expert users.

A. Citation-based Similarity Visualization

Our proposed citation exploration view for recommended papers uses interactive circle graphs to show the instances of shared citations within the user-selected reference document (visualized in gray) and a paper chosen from the set of paper recommendations (visualized on a white background).

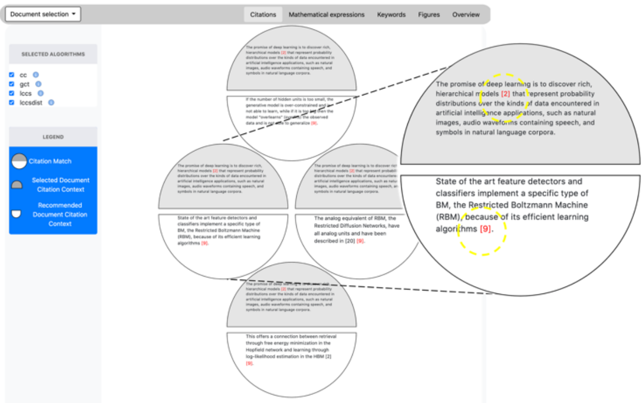

Figure 12. Detailed feature visualization view for academic citations.

Individual shared citations or citation groups are highlighted in red and placed in context by displaying their citing sentences, or *citances* [25]. Each circle corresponds to the same academic work being cited by both papers. The user can hover their mouse over the shared citation, shown in yellow in Figure 12, to display the citation's metadata. Readers can quickly scroll through all recommended papers that they have previously starred, and thus added to their collected documents (cf. Fig. 10). Seeing the similarity among the works cited and the individual citation contexts at a glance can help readers further decide which documents they want to read in-depth. Our detailed citation-based similarity examination view additionally supports transparency by allowing users to select the citation matching algorithm to be used from a left-hand panel. Explanations for the different citation-based recommendation approaches are expanded if the user clicks on the corresponding blue info icons (left panel in Fig. 12).

B. Keyword-based Similarity Visualization

The idea underlying the keyword-based view is to quickly gain an overview of both shared and unique keywords within the recommended set of papers being explored. The respective shared usage of keywords and their usage frequency are visualized by clustering keywords and using a color scale. Red hues visualize the most frequently shared terms, and blue hues indicate unique terms, while the size of words indicates their frequency within the papers being examined.

Figure 13. Keyword-based similarity visualization for comparing recommended papers in RecVis.

We chose this arrangement for visualizing keyword-based similarity among the recommended papers because it allows readers to identify the overlapping concepts (red) and the concepts unique to either paper (blue) without having to first skim the article text manually. As a result of comparing several recommended papers in this manner, readers can eliminate the papers that lack certain keywords of interest, or that frequently contain terms they do not find relevant. Thus, users can more effectively refine their selection of bookmarked papers for further examination or detailed reading.

C. Figure-based Similarity Visualization

In the figure-based similarity exploration view, researchers can quickly navigate between the figures in the recommended academic papers and filter the figures according to their similarity scores. If users wish to visually explore *all* figures present in the recommended paper, they can move the sensitivity threshold slider to zero, thus making all figures visible even in the absence of a detected feature-based similarity.

Since figures, images, and charts are already a compressed format for communicating an idea or data, the intuitive method for visualizing them remains the simple side-by-side comparison. In the RecVis prototype, the figures are shown together with their captions and their surrounding text, i.e. the paragraph above and below the extracted figure, since the text surrounding figures typically contains valuable contextual information relevant to the figure.

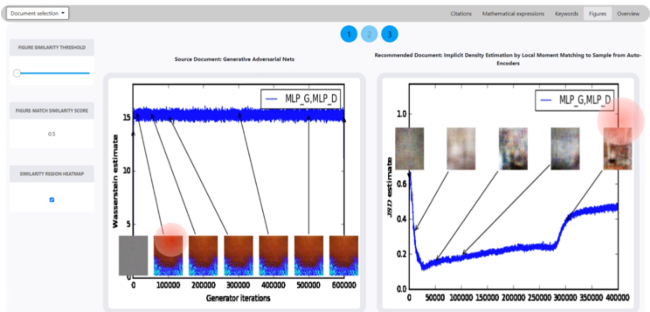

Figure 14. Figure-based similarity exploration view in RecVis.

In Fig. 14, the text surrounding the figures is not shown for the sake of the figures remaining slightly more legible in the layout of this publication. To increase the explainability of the recommendations, the user can activate an optional heatmap overlay to highlight the features or regions that were identified as having contributed the most to the similarity score. Together with the user-configurable sensitivity thresholds, these options for customization contribute to recommendation transparency.

D. Formula-based Similarity Visualization

The mathematical formula exploration view also uses a side-by-side comparison to visualize any formulae that have been extracted from the set of recommended research papers. The left panel includes a slider to adjust the threshold for displaying formulae based on their similarity score. By setting the slider to zero, users can once again browse all formulae or mathematical expressions found in the documents, even if they do not meet the minimum similarity requirements, i.e. if the formulae contain no matches of semantically equivalent expressions. Figure 15 shows the detailed exploration view for the mathematical formula-based exploration in the *RecVis* paper recommendation prototype. Semantically identical identifiers are highlighted in green, in this example the x, $=$, and the $\pm$ operator. Mathematical identifiers with high potential semantic similarity are highlighted in yellow. Additionally, the text surrounding the formulae in the full text of the documents is displayed, since this text is most likely to contain useful definitions or explanations.

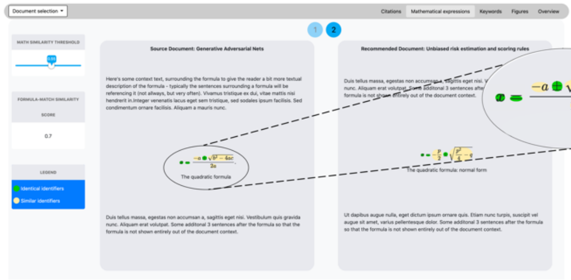

Figure 15. Mathematical formula similarity visualization in RecVis.

Note that for the mathematical formula exploration view, the RecVis prototype was forced to rely on manually generated formula data, since the document collection we use in RecVis did not yet contain sufficient mathematical expressions or formulae for the purpose of our evaluation. In upcoming work, we already plan to expand our recommendation corpus to include a subset of arXiv, which will contain the required math, statistics, and electrical engineering papers.

IV. EVALUATION

To evaluate the feasibility and user experience for detailed feature visualizations in the research paper recommendation context, we conducted an expert study with 10 participants. Our participants encompassed seven graduate students and three Ph.D. researchers from the University of Konstanz and the University of Wuppertal in Germany. Their profiles matched the 'expert' role, since all participants specified that they used academic paper search or recommendation systems for their work at least once per week, on average. Furthermore, our participants' academic backgrounds were rooted in the STEM fields (engineering, computer science, mathematics, and biology) thus allowing us to identify potential issues with visualizations that may be intuitive for a certain STEM field, but not for researchers from another background.

A. Data Collection

Data was collected in-person in early 2020 and with the help of a facilitator who interacted with each participant in a controlled setting on campus. The RecVis interface was shown on a laptop dedicated to usability testing purposes that was connected to a large, curved monitor. Each session lasted approx. 60 minutes per participant, on average. The data was collected in four ways. First, a think-aloud protocol was used throughout the study to facilitate the collection of relevant feedback from participants, which was audio recorded with the consent of participants. Second, the participants were given a set of 14 pre-defined tasks indicative of different information needs while interacting with the recommendation interface. We observed the ability of participants to perform the task accurately and retrieve the specified information. We categorized performance into successful task completion if the entire task was completed accurately, and task failure if a participant was unable to answer correctly, needed assistance, or answered only partially. Third, immediately following the task-based evaluation, participants were asked to fill out a customized survey addressing the ease of use and the perceived usefulness of different aspects of the visualizations with which they had just interacted. We used a 5-point Likert scale to measure the subjective satisfaction for the different views. Forth, and finally, our participants completed the standardized Post-Study System Usability Questionnaire (PSSUQ) [26] to measure user experience. PSSUQ is a good fit for our purposes, since the questionnaire reports on three categories 'system usefulness', 'information quality' and interface quality'. We do not use the recommendation system evaluation framework *ResQue* [27] in this context, because the objective of the evaluation presented in this paper is to test the *feasibility of feature-based exploration views in the research paper recommendation scenario*. Given this focus, our aim is to measure user satisfaction and the potential added value of the interactive interface, rather than recommendation accuracy, novelty, or diversity.

B. Results and Discussion

The results of this study help us answer whether the visualization of unique feature-based similarity can improve the research paper recommendation and exploration experience. Successful task completion rate across all 14 tasks and all participants was 95% on average. Completion rates ranged from an average of 70% for Task 1, to 100% for nine out of the 14 tasks. We found these results to be encouraging when keeping in mind that none of the specialized information exploration tasks (listed in Table 2) would have been supported by the existing visualization interfaces of paper recommendation systems. Instead, users would have been required to perform a time-consuming detailed reading and a manual comparison of the recommended papers. Table 2 lists all tasks posed in the study and their average completion rates among the participants.

Task 4, 5 and 14 had completion rates of 90%, and Task 9 had a completion rate of 80%. The poorest performing participant managed a completion rate of 86% and the best performing participants a completion rate of 100%. The task-based evaluation indicates that the interactive exploration and visualization of feature-based relevance we propose lead to good performance in successfully answering a range of specialized information needs regarding feature-based similarity.

In the following, we discuss the results of the user satisfaction ratings for the individual feature-based visualizations. User-perceived satisfaction was rated on a 5-point Likert scale, where the scores ranged from 'strongly disagree' (1) to 'strongly agree' (5). These questionnaire statements were aimed at spotting potential issues of the prototype after participants had completed the assigned task. Fig. 16, on the following page, uses diverging bar charts to show participants' perceived satisfaction with the individual feature exploration views. Additionally, participants rated their satisfaction with the user interface and the overview visualization.

Among the feature visualization categories, the lowest performing view was the keyword-based view, which averaged scores of 4.0, 3.9, and 4.1, respectively, for the three questions shown in Fig. 16. The formula-based visualization and the citation-based visualization were rated very positively with average Likert scores across all statements and all participants of 4.68 and 4.25, respectively.

In the general statements, participants were asked if dividing the similarity-features into separate views gave them a better understanding of the recommendation process and whether it helped them to focus on the task at hand. All participants either agreed or strongly agreed with these statements. Thus, we conclude that the visualization of individual features contained in STEM literature can improve the user experience for specialized information retrieval needs and for researchers well versed in the use of academic search and recommender systems.

TABLE II. STUDY TASKS

Task ID	***Feature-based view***	***IR Task Description***	***Successful completion rate (n = 10)***
1	*Citation exploration*	Identify the total number of citation-matches between the two inspected documents.	.70
2		Zoom in on a citation to make reading the contents easier.	1
3		Only display citation matches retrieved using the citation chunking (cc) algorithm	1
4	*Figure exploration*	Identify the total number of figure-matches between the two articles being compared.	.90
5		Identify the figure-match with the highest similarity score.	.90
6		In the figure view, enable the heatmap overlay.	1
7	*Mathematical formula exploration*	Filter the formula-matches to only show mathematical formulas with a similarity score of 0.7 or higher.	1
8		Identify the total number of formula-matches with a similarity of 0.7 or higher.	1
9	*Keyword-based exploration*	Identify three of the most common keywords that appear in the source document, but not in the recommended document.	.80
10		Identify three of the most common words that appear in the recommended, but not in the source document.	1
11		Identify three of the most common words used in both documents (there might be more than 3, simply choose 3).	1
12	*Global view & navigation*	Choose a different document that you want to compare in detail against the source document.	1
13		For a document of your choice, identify the feature which has contributed most to the overall similarity.	1
14		For each feature of the inspected document, identify the algorithm that provided the most results.	.90

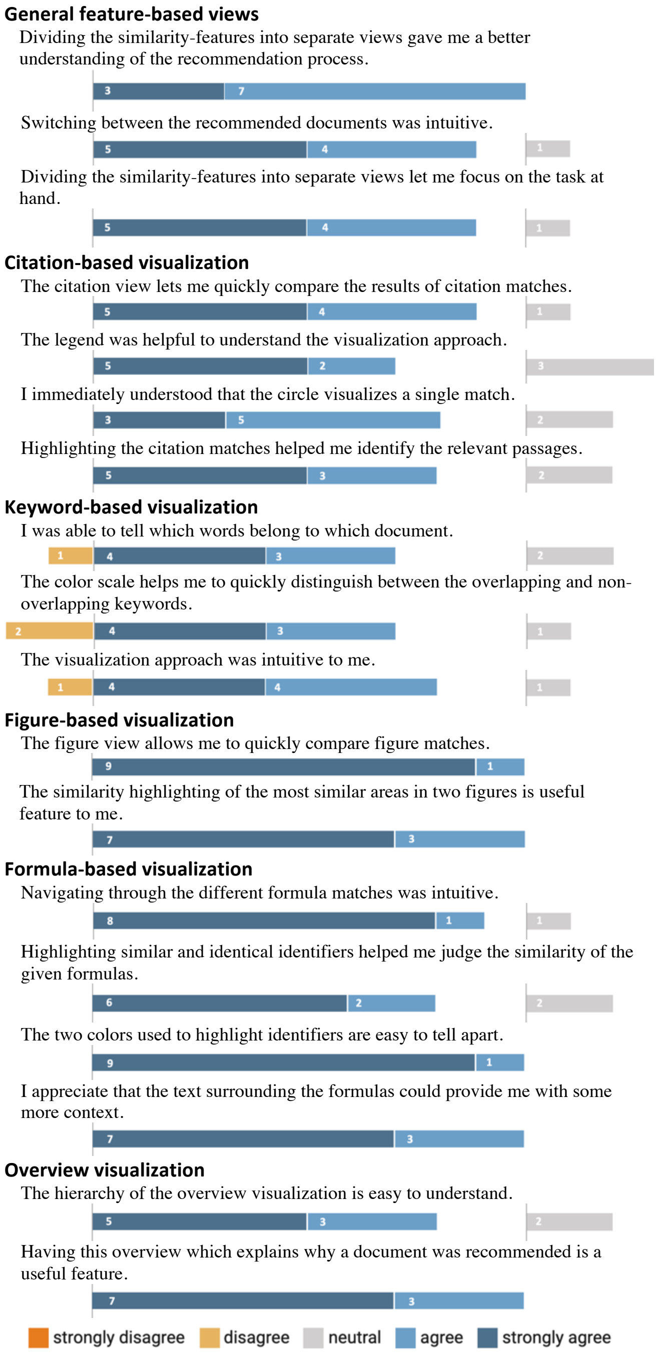

Figure 16. Frequency of ratings for individual feature exploration views.

Finally, to enable comparability with future studies and other interactive or exploratory recommendation systems, our participants completed the standardized PSSUQ. The PSSUQ consists of 16 questions, grouped into three sub-scores that reflect the following categories: *system usefulness*, *information quality*, and *interface quality*. Questions are rated on a scale from 1 – 7, where 1 is 'strongly agree' and 7 is 'strongly disagree'. We had to exclude questions 7 – 9 (which are related to error messages and recovery from mistakes) from the *information quality* score portion of the questionnaire, since three participants marked these questions as 'not applicable' because they experienced no error scenarios while interacting with the system. Information quality is thus calculated using three instead of six questions, and the overall score is based on 13 instead of 16 questions. The *system usability* portion of the questionnaire scored highest with an average score of 1.78, while the remaining three *information quality* questions scored the lowest, although still satisfactory, with an average score of 2.23. The average score for *interface quality* was 1.85, and the overall score was a favorable 1.91. As a result of our evaluation, we have already made some design improvements, e.g., changed the way explanations are displayed in the keyword-based exploration view and how citation-matches are visualized in the citation exploration view. We will evaluate these changes in an upcoming online study with more participants and by utilizing a between-subject design, where we will additionally compare the feature-based exploration views to a baseline academic paper recommendation interface.

A limitation of this in-person study is that it only focused on general usability questions and that the task-based feasibility evaluation was for a set of artificially defined information seeking objectives. In a future study, we plan to recruit more domain experts, specifically from mathematics and from the biomedical domain, to help us further customize the mathematical formula and figure visualizations to the specialized needs of domain experts, whose needs we are likely to have missed in this first evaluation. We are currently expanding the paper recommendation collection and plan to include research papers from the domains of the experts partaking in the upcoming evaluation, so they will be able to better judge the quality of the feature visualizations, as well as the recommendation quality.

At this point, we wish to reiterate that we are not arguing for feature-based similarity to become the central basis to satisfy the need for visualizations in article recommendation systems in general. Instead, we see feature-based visualizations as a valuable supplementary view to address the highly specialized information needs of domain experts, or to further narrow down a set of literature recommendations. While the list-based ranking of similar articles will remain at the core of academic recommender systems due to its simplicity and space-saving format, our system and the proposed feature visualizations can augment existing recommendation interfaces to support those researchers with more specialized search or recommendation needs relating to the mathematical formulas, charts and figures or citations contained in academic publications.

In summary, our expert users agreed that dividing the feature-based similarity within publications into separate views gave them a better understanding of the feature's role in the recommendations made. Additionally, such a view could help researchers identify the papers they intend to read in their entirety more quickly. The task-based assessment showed that our interactive visualization of feature-based similarity lets researchers perform specialized information retrieval tasks that are not supported by traditional recommendation interfaces.

V. Conclusion

We surveyed existing research on interactive visualization approaches in recommender systems. Due to the scarcity of research in this field, in particular, for the academic literature domain, we additionally surveyed feature-based visualizations for academic papers as proposed in the near-duplicate detection domain. We found that current visualization techniques do not offer the required resolution to address specialized information seeking needs in the paper recommendation use case. Visualizations for non-textual features contained in research papers (i.e., citation patterns, figures, charts and images, and mathematical formulas) have thus far only been considered in the near-duplicate detection domain. No feature-based visualizations exist in the paper recommendation domain to support the more specialized literature search and exploration needs of researchers. For example, our proposed feature-based exploration views are especially valuable to STEM researchers, who may wish to discover the instances of mathematical formula similarity, chart and figure-based similarity, or the shared citations present in a set of literature recommendations. Being able to quickly discover such forms of similarity is valuable when seeking literature recommendations that are relevant to a certain mathematical proof or that apply a specific formula to a problem, that discuss the same references, or that cite the same figures or tables. Our work contributes to filling this research gap by proposing an interactive feature exploration visualization for the academic literature recommendation use case. The proposed visualizations of *RecVis* allow comparing documents side-by-side on the basis of their contained features, while supporting filtering and navigation within the entire set of recommended documents, which we display in a force-direct graph layout. The results of an evaluation with expert users show that the interactive interface we propose allows a quick identification of features of the interest, which enable users to answer specialized information needs that existing recommendation exploration interfaces cannot support. The recommender system and proposed visualization methods implemented in *RecVis* give actionable insight to practitioners building and improving academic literature recommender systems.

Acknowledgment

This work was funded by the Carl Zeiss Foundation and the German Research Foundation (DFG) Grant No.: GI 1259/1.

References

[1] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: a literature survey," *International Journal on Digital Libraries*, no. 17, pp. 305–338, Nov. 2016, Available: https://doi.org/10.1007/s00799-015-0156-0.

[2] C.-H. Tsai and P. Brusilovsky, "Beyond the ranked list: User-driven exploration and diversification of social recommendation," in *23rd International Conference on Intelligent User Interfaces*, New York, NY, USA, Mar. 2018, pp. 239–250. Available: https://doi.org/10.1145/3172944.3172959

[3] C. Breitinger, B. Kolcu, M. Meuschke, N. Meuschke, and B. Gipp, "Supporting the exploration of semantic features in academic literature using graph-based visualizations," presented at the ACM/IEEE Joint Conference on Digital Libraries (JCDL), Jul. 2020.

[4] J. Beel, A. Aizawa, C. Breitinger, and B. Gipp, "Mr. DLib: Recommendations-as-a-Service (RaaS) for academia," presented at the Proceedings of the Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2017.

[5] C. Tenopir, L. Christian, and J. Kaufman, "Seeking, reading, and use of scholarly articles: An international study of perceptions and behavior of researchers," *Publications*, vol. 7, no. 1, Art. no. 1, Mar. 2019, doi: 10.3390/publications7010018.

[6] K. Verbert, D. Parra, P. Brusilovsky, and E. Duval, "Visualizing recommendations to support exploration, transparency and controllability," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, Santa Monica, CA, USA, Mar. 2013, pp. 351–362. doi: 10.1145/2449396.2449442.

[7] H. Cramer *et al.*, "The effects of transparency on trust in and acceptance of a content-based art recommender," *User Model User-Adap Inter*, vol. 18, no. 5, p. 455, Aug. 2008, doi: 10.1007/s11257-008-9051-3.

[8] I. Nunes and D. Jannach, "A systematic review and taxonomy of explanations in decision support and recommender systems," *User Model User-Adap Inter*, vol. 27, no. 3–5, pp. 393–444, Dec. 2017, doi: 10.1007/s11257-017-9195-0.

[9] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives," *FNT in Information Retrieval*, vol. 14, no. 1, pp. 1–101, 2020, doi: 10.1561/1500000066.

[10] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: evaluating recommender systems by coverage and serendipity," in *Proceedings of the fourth ACM conference on Recommender Systems*, Barcelona, Spain, Sep. 2010, pp. 257–260. doi: 10.1145/1864708.1864761.

[11] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work*, Philadelphia, Pennsylvania, USA, Dec. 2000, pp. 241–250. doi: 10.1145/358916.358995.

[12] M. Jugovac and D. Jannach, "Interacting with recommenders - overview and research directions," *ACM Trans. Interact. Intell. Syst.*, vol. 7, no. 3, p. 10:1-10:46, Sep. 2017, doi: 10.1145/3001837.

[13] B. Gipp, J. Beel, and C. Hentschel, "Scienstein: A research paper recommender system," presented at the *Proceedings of the International Conference on Emerging Trends in Computing* (ICETiC'09), Virudhunagar, India, Jan. 2009.

[14] "Connected Papers," *(webpage)*. url: https://www.connectedpapers.com/ (accessed Jun. 01, 2021).

[15] D. Parra, P. Brusilovsky, and C. Trattner, "See what you want to see: visual user-driven approach for hybrid recommendation," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, New York, NY, USA, Feb. 2014, pp. 235–240. doi: 10.1145/2557500.2557542.

[16] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "TasteWeights: a visual interactive hybrid recommender system," in *Proceedings of the sixth ACM conference on Recommender systems*, New York, NY, USA, Sep. 2012, pp. 35–42. doi: 10.1145/2365952.2365964.

[17] J. O'Donovan, B. Smyth, B. Gretarsson, S. Bostandjiev, and T. Höllerer, "PeerChooser: visual interactive recommendation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2008, pp. 1085–1088. doi: 10.1145/1357054.1357222.

[18] L. Gou, F. You, J. Guo, L. Wu, and X. (Luke) Zhang, "SFViz: interest-based friends exploration and recommendation in social networks," in *Proceedings of the 2011 Visual Information Communication - International Symposium*, New York, NY, USA, Aug. 2011, pp. 1–10. doi: 10.1145/2016656.2016671.

[19] B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, and T. Höllerer, "SmallWorlds: Visualizing social recommendations," *Computer Graphics Forum*, vol. 29, no. 3, pp. 833–842, 2010, doi: 10.1111/j.1467-8659.2009.01679.x.

[20] N. Meuschke, V. Stange, M. Schubotz, and B. Gipp, "HyPlag: A hybrid approach to academic plagiarism detection," in *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Ann Arbor, MI, USA, Jun. 2018, pp. 1321–1324. doi: 10.1145/3209978.3210177.

[21] T. Foltýnek, N. Meuschke, and B. Gipp, "Academic Plagiarism Detection: A Systematic Literature Review," *ACM Comput. Surv.*, vol. 52, no. 6, p. 112:1-112:42, Oct. 2019, doi: 10.1145/3345317.

[22] N. Meuschke, B. Gipp, and C. Breitinger, "CitePlag: A citation-based plagiarism detection system prototype," in *Proceedings of the 5th International Plagiarism Conference*, Newcastle upon Tyne, UK, Jul. 2012.

[23] B. Gipp, "Citation-based Plagiarism Detection - Detecting Disguised and Cross-language Plagiarism using Citation Pattern Analysis," Springer Vieweg Research, 2014. doi: 10.1007/978-3-658-06394-8.

[24] P. Riehmann, M. Potthast, B. Stein, and B. Froehlich, "Visual assessment of alleged plagiarism cases," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 61–70, 2015.

[25] P. I. Nakov, A. S. Schwartz, and M. A. Hearst, "Citances: citation sentences for semantic analysis of bioscience text," in Proceedings of the SIGIR'04 workshop on Search and Discovery in Bioinformatics, Sheffield, UK, 2004.

[26] J. R. Lewis, "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," *International Journal of Human–Computer Interaction*, vol. 7, no. 1, pp. 57–78, Jan. 1995, doi: 10.1080/10447319509526110.

[27] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, Chicago, Illinois, USA, Oct. 2011, pp. 157–164. doi: 10.1145/2043932.2043962.